

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 355 149 A2

(12)

EUROPEAN PATENT APPLICATION

(43) Date of publication:
22.10.2003 Bulletin 2003/43

(51) Int Cl.7: **G01N 33/48**, C07H 21/04,
C12Q 1/68

(21) Application number: **03252022.3**

(22) Date of filing: **31.03.2003**

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PT RO SE SI SK TR
Designated Extension States:
AL LT LV MK

(72) Inventor: **Wang, Yixin**
San Diego, CA 92130 (US)

(74) Representative: **Mercer, Christopher Paul et al**
Carpmaels & Ransford
43, Bloomsbury Square
London WC1A 2RA (GB)

(30) Priority: **29.03.2002 US 368798 P**

(71) Applicant: **Ortho-Clinical Diagnostics, Inc.**
Rochester, NY 14626-5101 (US)

(54) **Assessing colorectal cancer**

(57) A method of assessing the presence or absence of colorectal cancer or the likely condition of a person believed to have colorectal cancer is conducted by analyzing the expression of a group of genes. Gene

expression profiles in a variety of medium such as microarrays are included as are kits that contain them.

EP 1 355 149 A2

Description**BACKGROUND**

5 [0001] This application claims the benefit of U.S. Provisional Application No. 60/368,798 filed on March 29, 2002.

[0002] This invention relates to diagnostics and prognostics for colorectal cancer based on the gene expression profiles of biological samples.

10 [0003] Colorectal cancer is a heterogeneous disease, consisting of tumors thought to emerge through three major molecular mechanisms: 1) mutations in the adenomatous polyposis coli (APC) gene, or the β -catenin gene, combined with chromosomal instability, 2) mutations in DNA mismatch repair genes, such as MLH1, MSH2, PMS1, PMS2 and MSH6, associated with microsatellite instability and mutations in genes containing short repeats, and 3) gene silencing induced by hypermethylation of the promoter regions of tumor suppressor genes. The genetic complement of individual colorectal cancers is likely to include different combinations of genetic instability, specific mutations, and gene silencing. Chromosomal instability (CIN) is a common feature of cancers in general. It implies an aneuploid phenotype, in which 15 whole chromosomes or large parts of them are being lost or gained. Microsomal instability (MIN) is found in diploid tumors with an increased mutation rate in short repeats. Both forms of genetic instability are common in colorectal cancer.

20 [0004] Colorectal cancers thus have complex origins and involve a number of interactions in different biological pathways. Serum markers, histological, and cytological examinations historically used to assist in providing diagnostic, prognostic, or therapy monitoring decisions often do not have desired reliability. Likewise, while use of a single genetic marker (e.g., increased expression of a particular gene) may be beneficial, the diversity of the cancers make it more likely that a portfolio of genetic markers is the best approach.

SUMMARY OF THE INVENTION

25 [0005] The invention is a method of assessing the presence or absence of colorectal cancer or the likely condition of a person believed to have colorectal cancer. In the method, a gene expression profile of a patient sample is analyzed to determine whether a patient has a colorectal cancer, whether a patient does not have colorectal cancer, whether a patient is likely to get colorectal cancer, or the response to treatment of a patient being treated for colorectal cancer.

30 [0006] Articles used in practicing the methods are also an aspect of the invention. Such articles include gene expression profiles or representations of them that are fixed in machine-readable media such as computer readable media.

[0007] Articles used to identify gene expression profiles can also include substrates or surfaces, such as microarrays, to capture and/or indicate the presence, absence, or degree of gene expression.

DETAILED DESCRIPTION

40 [0008] The mere presence or absence of particular nucleic acid sequences in a tissue sample has only rarely been found to have diagnostic or prognostic value. Information about the expression of various proteins, peptides or mRNA, on the other hand, is increasingly viewed as important. The mere presence of nucleic acid sequences having the potential to express proteins, peptides, or mRNA (such sequences referred to as "genes") within the genome by itself is not determinative of whether a protein, peptide, or mRNA is expressed in a given cell. Whether or not a given gene capable of expressing proteins, peptides, or mRNA does so and to what extent such expression occurs, if at all, is determined by a variety of complex factors. Irrespective of difficulties in understanding and assessing these factors, 45 assaying gene expression can provide useful information about the occurrence of important events such as tumorigenesis, metastasis, apoptosis, and other clinically relevant phenomena. Relative indications of the degree to which genes are active or inactive can be found in gene expression profiles. The gene expression profiles of this invention are used to diagnose and treat patients for colorectal cancer.

50 [0009] Sample preparation requires the collection of patient samples. Patient samples used in the inventive method are those that are suspected of containing diseased cells such as epithelial cells taken from a colon sample or from surgical margins. One useful technique for obtaining suspect samples is Laser Capture Microdissection (LCM). LCM technology provides a way to select the cells to be studied, minimizing variability caused by cell type heterogeneity. Consequently, moderate or small changes in gene expression between normal and cancerous cells can be readily detected. In a preferred method, the samples comprise circulating epithelial cells extracted from peripheral blood. 55 These can be obtained according to a number of methods but the most preferred method is the magnetic separation technique described in U.S. Patent 6,136,182 assigned to Immunivest Corp which is incorporated herein by reference. Once the sample containing the cells of interest has been obtained, RNA is extracted and amplified and a gene expression profile is obtained, preferably via micro-array, for genes in the appropriate portfolios.

[0010] Preferred methods for establishing gene expression profiles include determining the amount of RNA that is produced by a gene that can code for a protein or peptide. This is accomplished by reverse transcriptase PCR (RT-PCR), competitive RT-PCR, real time RT-PCR, differential display RT-PCR, Northern Blot analysis and other related tests. While it is possible to conduct these techniques using individual PCR reactions, it is best to amplify complementary DNA (cDNA) or complementary RNA (crRNA) produced from mRNA and analyze it via microarray. A number of different array configurations and methods for their production are known to those of skill in the art and are described in U.S. Patents such as: 5,445,934; 5,532,128; 5,556,752; 5,242,974; 5,384,261; 5,405,783; 5,412,087; 5,424,186; 5,429,807; 5,436,327; 5,472,672; 5,527,681; 5,529,756; 5,545,531; 5,554,501; 5,561,071; 5,571,639; 5,593,839; 5,599,695; 5,624,711; 5,658,734; and 5,700,637; the disclosures of which are incorporated herein by reference.

[0011] Microarray technology allows for the measurement of the steady-state mRNA level of thousands of genes simultaneously thereby presenting a powerful tool for identifying effects such as the onset, arrest, or modulation of uncontrolled cell proliferation. Two microarray technologies are currently in wide use. The first are cDNA arrays and the second are oligonucleotide arrays. Although differences exist in the construction of these chips, essentially all downstream data analysis and output are the same. The product of these analyses are typically measurements of the intensity of the signal received from a labeled probe used to detect a cDNA sequence from the sample that hybridizes to a nucleic acid sequence at a known location on the microarray. Typically, the intensity of the signal is proportional to the quantity of cDNA, and thus mRNA, expressed in the sample cells. A large number of such techniques are available and useful. Preferred methods for determining gene expression can be found in US Patents 6,271,002 to Linsley, et al.; 6,218,122 to Friend, et al.; 6,218,114 to Peck, et al.; and 6,004,755 to Wang, et al., the disclosure of each of which is incorporated herein by reference.

[0012] Analysis of the expression levels is conducted by comparing such intensities. This is best done by generating a ratio matrix of the expression intensities of genes in a test sample versus those in a control sample. For instance, the gene expression intensities from a diseased tissue can be compared with the expression intensities generated from normal tissue of the same type (e.g., diseased colon tissue sample vs. normal colon tissue sample). A ratio of these expression intensities indicates the fold-change in gene expression between the test and control samples.

[0013] Gene expression profiles can also be displayed in a number of ways. The most common method is to arrange a raw fluorescence intensities or ratio matrix into a graphical dendrogram where columns indicate test samples and rows indicate genes. The data is arranged so genes that have similar expression profiles are proximal to each other. The expression ratio for each gene is visualized as a color. For example, a ratio less than one (indicating down-regulation) may appear in the blue portion of the spectrum while a ratio greater than one (indicating up-regulation) may appear as a color in the red portion of the spectrum. Commercially available computer software programs are available to display such data including "GENESPRING" from Silicon Genetics, Inc. and "DISCOVERY" and "INFER" software from Partek, Inc.

[0014] Modulated genes used in the methods of the invention are shown in Table 1. The genes that are differentially expressed are shown as being either up regulated or down regulated in diseased cells. Up regulation and down regulation are relative terms meaning that a detectable difference (beyond the contribution of noise in the system used to measure it) is found in the amount of expression of the genes relative to some baseline. In this case, the baseline is the measured gene expression of a normal cell. The genes of interest in the diseased cells are then either up regulated or down regulated relative to the baseline level using the same measurement method. Diseased, in this context, refers to an alteration of the state of a body that interrupts or disturbs, or has the potential to disturb, proper performance of bodily functions as occurs with the uncontrolled proliferation of cells. Someone is diagnosed with a disease when some aspect of that person's genotype or phenotype is consistent with the presence of the disease. However, the act of conducting a diagnosis or prognosis includes the determination disease/status issues such as therapy monitoring. In therapy monitoring, clinical judgments are made regarding the effect of a given course of therapy by comparing the expression of genes over time to determine whether the gene expression profiles have changed or are changing to patterns more consistent with normal tissue.

[0015] Preferably, levels of up and down regulation are distinguished based on fold changes of the intensity measurements of hybridized microarray probes. A 2.0 fold difference is preferred for making such distinctions or a p-value less than .05. That is, before a gene is said to be differentially expressed in diseased versus normal cells, the diseased cell is found to yield at least 2 more, or 2 times less intensity than the normal cells. The greater the fold difference, the more preferred is use of the gene as a diagnostic. Genes selected for the gene expression profiles of the instant invention have expression levels that result in the generation of a signal that is distinguishable from those of the normal or non-modulated genes by an amount that exceeds background using clinical laboratory instrumentation.

[0016] Statistical values can be used to confidently distinguish modulated from non-modulated genes and noise. Statistical tests find the genes most significantly different between diverse groups of samples. The Student's t-test is an example of a robust statistical test that can be used to find significant differences between two groups. The lower the p-value, the more compelling the evidence that the gene is showing a difference between the different groups. Nevertheless, since microarrays measure more than one gene at a time, tens of thousands of statistical tests may be

asked at one time. Because of this, there is likelihood to see small p-values just by chance and adjustments for this using a Sidak correction as well as a randomization/permutation experiment can be made. A p-value less than .05 by the t-test is evidence that the gene is significantly different. More compelling evidence is a p-value less than .05 after the Sidak correct is factored in. For a large number of samples in each group, a p-value less than 0.05 after the randomization/permutation test is the most compelling evidence of a significant difference.

[0017] Another parameter that can be used to select genes that generate a signal that is greater than that of the non-modulated gene or noise is the use of a measurement of absolute signal difference. Preferably, the signal generated by the modulated gene expression is at least 20% different than those of the normal or non-modulated gene (on an absolute basis). It is even more preferred that such genes produce expression patterns that are at least 30% different than those of normal or non-modulated genes.

[0018] Genes can be grouped so that information obtained about the set of genes in the group provides a sound basis for making a clinically relevant judgment such as a diagnosis, prognosis, or treatment choice. These sets of genes make up the portfolios of the invention. In this case, the judgments supported by the portfolios involve colorectal cancer. Portfolios of gene expression profiles can be comprised of combinations of genes described in Example 3. As with most diagnostic markers, it is often desirable to use the fewest number of markers sufficient to make a correct medical judgment. This prevents a delay in treatment pending further analysis as well inappropriate use of time and resources. In this case, such a minimal portfolio can be comprised of a combination of genes from Example 4.

[0019] Preferably, portfolios are established such that the combination of genes in the portfolio exhibit improved sensitivity and specificity relative to individual genes or randomly selected combinations of genes. In the context of the instant invention, the sensitivity of the portfolio can be reflected in the fold differences exhibited by a gene's expression in the diseased state relative to the normal state. Specificity can be reflected in statistical measurements of the correlation of the signaling of gene expression with the condition of interest. For example, standard deviation can be used as such a measurement. In considering a group of genes for inclusion in a portfolio, a small standard deviation in expression measurements correlates with greater specificity. Other measurements of variation such as correlation coefficients can also be used in this capacity. The most preferred method of establishing gene expression portfolios is through the use of optimization algorithms such as the mean variance algorithm widely used in establishing stock portfolios. This method is described in detail in the co-pending patent application entitled "Portfolio Selection" by Tim Jatkoe, et. al., of equal date hereto. Essentially, the method calls for the establishment of a set of inputs (stocks in financial applications, expression as measured by intensity here) that will optimize the return (e.g., signal that is generated) one receives for using it while minimizing the variability of the return. Many commercial software programs are available to conduct such operations. "Wagner Associates Mean-Variance Optimization Application", referred to as "Wagner Software" throughout this specification, is preferred. This software uses functions from the "Wagner Associates Mean-Variance Optimization Library" to determine an efficient frontier and optimal portfolios in the Markowitz sense is preferred.

[0020] Use of this type of software requires that microarray data be transformed so that it can be treated as an input in the way stock return and risk measurements are used when the software is used for its intended financial analysis purposes. For example, when Wagner Software is employed in conjunction with microarray intensity measurements the following data transformation method is employed.

[0021] Genes are first pre-selected by identifying those genes whose expression shows at least some minimal level of differentiation. The preferred pre-selection process is conducted as follows. A baseline class is selected. Typically, this will comprise genes from a population that does not have the condition of interest. For example, if one were interested in selecting a portfolio of genes that are diagnostic for breast cancer, samples from patients without breast cancer can be used to make the baseline class. Once the baseline class is selected, the arithmetic mean and standard deviation is calculated for the indicator of gene expression of each gene for baseline class samples. This indicator is typically the fluorescent intensity of a microarray reading. The statistical data computed is then used to calculate a baseline value of $(X \times \text{Standard Deviation} + \text{Mean})$ for each gene. This is the baseline reading for the gene from which all other samples will be compared. X is a stringency variable selected by the person formulating the portfolio. Higher values of X are more stringent than lower. Preferably, X is in the range of .5 to 3 with 2 to 3 being more preferred and 3 being most preferred.

[0022] Ratios between each experimental sample (those displaying the condition of interest) versus baseline readings are then calculated. The ratios are then transformed to base 10 logarithmic values for ease of data handling by the software. This enables down regulated genes to display negative values necessary for optimization according to the Markman mean-variance algorithm using the Wagner Software.

[0023] The preprocessed data comprising these transformed ratios are used as inputs in place of the asset return values that are normally used in the Wagner Software when it is used for financial analysis purposes.

[0024] Once an efficient frontier is formulated, an optimized portfolio is selected for a given input level (return) or variance that corresponds to a point on the frontier. These inputs or variances are the predetermined standards set by the person formulating the portfolio. Stated differently, one seeking the optimum portfolio determines an acceptable

input level (indicative of sensitivity) or a given level of variance (indicative of specificity) and selects the genes that lie along the efficient frontier that correspond to that input level or variance. The Wagner Software can select such genes when an input level or variance is selected. It can also assign a weight to each gene in the portfolio as it would for a stock in a stock portfolio.

[0025] Determining whether a sample has the condition for which the portfolio is diagnostic can be conducted by comparing the expression of the genes in the portfolio for the patient sample with calculated values of differentially expressed genes used to establish the portfolio. Preferably, a portfolio value is first generated by summing the multiples of the intensity value of each gene in the portfolio by the weight assigned to that gene in the portfolio selection process. A boundary value is then calculated by $(Y \times \text{standard deviation} + \text{mean of the portfolio value for baseline groups})$ where Y is a stringency value having the same meaning as X described above. A sample having a portfolio value greater than the portfolio value of the baseline class is then classified as having the condition. If desired, this process can be conducted iteratively in accordance with well known statistical methods for improving confidence levels.

[0026] Optionally one can reiterate this process until best prediction accuracy is obtained.

[0027] The process of portfolio selection and characterization of an unknown is summarized as follows:

1. Choose baseline class.
2. Calculate mean, and standard deviation of each gene for baseline class samples.
3. Calculate $(X \times \text{Standard Deviation} + \text{Mean})$ for each gene. This is the baseline reading from which all other samples will be compared. X is a stringency variable with higher values of X being more stringent than lower.
4. Calculate ratio between each Experimental sample versus baseline reading calculated in step 3.
5. Transform ratios such that ratios less than 1 are negative (eg.using Log base 10). (Down regulated genes now correctly have negative values necessary for MV optimization).
6. These transformed ratios are used as inputs in place of the asset returns that are normally used in the software application.
7. The software will plot the efficient frontier and return an optimized portfolio at any point along the efficient frontier.
8. Choose a desired return or variance on the efficient frontier.
9. Calculate the Portfolio's Value for each sample by summing the multiples of each gene's intensity value by the weight generated by the portfolio selection algorithm.
10. Calculate a boundary value by adding the mean Portfolio Value for Baseline groups to the multiple of Y and the Standard Deviation of the Baseline's Portfolio Values. Values greater than this boundary value shall be classified as the Experimental Class.
11. Optionally one can reiterate this process until best prediction accuracy is obtained.

[0028] Alternatively, genes can first be pre-selected by identifying those genes whose expression shows some minimal level of differentiation. The pre-selection in this alternative method is preferably based on a threshold given by

$$1 \leq \left| \frac{(\mu_t - \mu_n)}{(\sigma_t + \sigma_n)} \right|,$$

where μ_t is the mean of the subset known to possess the disease or condition, μ_n is the mean of the subset of normal samples, and $\sigma_t + \sigma_n$ represent the combined standard deviations. A signal to noise cutoff can also be used by pre-selecting the data according to a relationship such as

$$0.5 \leq \left| \frac{(\mu_t - MAX_n)}{(\sigma_t + \sigma_n)} \right|.$$

This ensures that genes that are pre-selected based on their differential modulation are differentiated in a clinically significant way. That is, above the noise level of instrumentation appropriate to the task of measuring the diagnostic parameters. For each marker pre-selected according to these criteria, a matrix is established in which columns represents samples, rows represent markers and each element is a normalized intensity measurement for the expression of that marker according to the relationship:

$$\left| \frac{(\mu_t - I)}{\mu_t} \right|$$

where I is the intensity measurement.

[0029] It is also possible to set additional boundary conditions to define the optimal portfolios. For example, portfolio size can be limited to a fixed range or number of markers. This can be done either by making data pre-selection criteria more stringent (e.g.,

$$.8 \leq \left| \frac{(\mu_I - MAX_n)}{(\sigma_I + \sigma_n)} \right|$$

instead of

$$0.5 \leq \left| \frac{(\mu_I - MAX_n)}{(\sigma_I + \sigma_n)} \right|$$

or by using programming features such as restricting portfolio size. One could, for example, set the boundary condition that the efficient frontier is to be selected from among only the most optimal 10 genes. One could also use all of the genes pre-selected for determining the efficient frontier and then limit the number of genes selected (e.g., no more than 10).

[0030] The process of selecting a portfolio can also include the application of heuristic rules. Preferably, such rules are formulated based on biology and an understanding of the technology used to produce clinical results. More preferably, they are applied to output from the optimization method. For example, the mean variance method of portfolio selection can be applied to microarray data for a number of genes differentially expressed in subjects with breast cancer. Output from the method would be an optimized set of genes that could include some genes that are expressed in peripheral blood as well as in diseased breast tissue. If sample used in the testing method are obtained from peripheral blood and certain genes differentially expressed in instances of breast cancer could also be differentially expressed in peripheral blood, then a heuristic rule can be applied in which a portfolio is selected from the efficient frontier excluding those that are differentially expressed in peripheral blood. Of course, the rule can be applied prior to the formation of the efficient frontier by, for example, applying the rule during data pre-selection.

[0031] Other heuristic rules can be applied that are not necessarily related to the biology in question. For example, one can apply the rule that only a given percentage of the portfolio can be represented by a particular gene or genes. Commercially available software such as the Wagner Software readily accommodates these types of heuristics. This can be useful, for example, when factors other than accuracy and precision (e.g., anticipated licensing fees) have an impact on the desirability of including one or more genes.

[0032] One method of the invention involves comparing gene expression profiles for various genes (or portfolios) to conduct diagnoses as described above. The gene expression profiles of each of the genes comprising the portfolio are fixed in a medium such as a computer readable medium. This can take a number of forms. For example, a table can be established into which the range of signals (e.g., intensity measurements) indicative of disease is input. Actual patient data can then be compared to the values in the table to determine whether the patient samples are normal or diseased. In a more sophisticated embodiment, patterns of the expression signals (e.g., fluorescent intensity) are recorded digitally or graphically. The gene expression patterns from the gene portfolios used in conjunction with patient samples are then compared to the expression patterns. Pattern comparison software can then be used to determine whether the patient samples have a pattern indicative of the disease in question. Of course, these comparisons can also be used to determine whether the patient results are normal. The expression profiles of the samples are then compared to the portfolio of a normal or control cell. If the sample expression patterns are consistent with the expression pattern for a colorectal cancer then (in the absence of countervailing medical considerations) the patient is diagnosed as positive for colorectal cancer. If the sample expression patterns are consistent with the expression pattern from the normal/control cell then the patient is diagnosed negative for colorectal cancer.

[0033] Numerous well known methods of pattern recognition are available. The following references provide some examples:

Weighted Voting: Golub, TR., Slonim, DK., Tamaya, P., Huard, C., Gaasenbeek, M., Mesirov, JP., Coller, H., Loh, L., Downing, JR., Caligiuri, MA., Bloomfield, CD., Lander, ES. *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. Science 286:531-537, 1999

Support Vector Machines: Su, AI., Welsh, JB., Sapinoso, LM., Kern, SG., Dimitrov, P., Lapp, H., Schultz, PG., Powell, SM., Moskaluk, CA., Frierson, HF. Jr., Hampton, GM. *Molecular classification of human carcinomas by use of gene expression signatures*. Cancer Research 61:7388-93, 2001 and

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, CH., Angelo, M., Ladd, C., Reich, M., Latu-

lippe, E., Mesirov, JP., Poggio, T., Gerald, W., Loda, M., Lander, ES., Gould, TR. *Multiclass cancer diagnosis using tumor gene expression signatures* Proceedings of the National Academy of Sciences of the USA 98:15149-15154, 2001

K-nearest Neighbors: Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, CH., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, JP., Poggio, T., Gerald, W., Loda, M., Lander, ES., Gould, TR. *Multiclass cancer diagnosis using tumor gene expression signatures* Proceedings of the National Academy of Sciences of the USA 98:15149-15154, 2001

Correlation Coefficients: van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002 Jan 31;415(6871):530-6.

[0034] The gene expression profiles of this invention can also be used in conjunction with other non-genetic diagnostic methods useful in cancer diagnosis, prognosis, or treatment monitoring. For example, in some circumstances it is beneficial to combine the diagnostic power of the gene expression based methods described above with data from conventional markers such as serum protein markers (e.g., carcinoembryonic antigen). A range of such markers exists including such analytes as CA19-9, CA 125, CK-BB, and Guanylyl Cyclase C. In one such method, blood is periodically taken from a treated patient and then subjected to an enzyme immunoassay for one of the serum markers described above. When the concentration of the marker suggests the return of tumors or failure of therapy, a sample source amenable to gene expression analysis is taken. Where a suspicious mass exists, a fine needle aspirate is taken and gene expression profiles of cells taken from the mass are then analyzed as described above. Alternatively, tissue samples may be taken from areas adjacent to the tissue from which a tumor was previously removed. This approach can be particularly useful when other testing produces ambiguous.

[0035] Combining the use of genetic markers with other diagnostics is most preferred when the reliability of the other diagnostic is suspect. For example, it is known that serum levels of CEA can be substantially affected by factors having nothing to do with a patient's cancer status. It can be beneficial to conduct a combination gene expression/CEA assay when a patient being monitored following treatment for colon cancer shows heightened levels of routine CEA assays.

[0036] Articles of this invention include representations of the gene expression profiles useful for treating, diagnosing, prognosticating, and otherwise assessing diseases. These profile representations are reduced to a medium that can be automatically read by a machine such as computer readable media (magnetic, optical, and the like). The articles can also include instructions for assessing the gene expression profiles in such media. For example, the articles may comprise a CD ROM having computer instructions for comparing gene expression profiles of the portfolios of genes described above. The articles may also have gene expression profiles digitally recorded therein so that they may be compared with gene expression data from patient samples. Alternatively, the profiles can be recorded in different representational format. A graphical recordation is one such format. Clustering algorithms such as those incorporated in "GENESPRING" and "DISCOVER" computer programs mentioned above can best assist in the visualization of such data.

[0037] Different types of articles of manufacture according to the invention are media or formatted assays used to reveal gene expression profiles. These can comprise, for example, microarrays in which sequence complements or probes are affixed to a matrix to which the sequences indicative of the genes of interest combine creating a readable determinant of their presence. Alternatively, articles according to the invention can be fashioned into reagent kits for conducting hybridization, amplification, and signal generation indicative of the level of expression of the genes of interest for detecting colorectal cancer.

[0038] Kits made according to the invention include formatted assays for determining the gene expression profiles. These can include all or some of the materials needed to conduct the assays such as reagents and instructions.

[0039] The invention is further illustrated by the following non-limiting examples.

Examples: Genes analyzed according to this invention are identified by reference to Gene ID Numbers in the GenBank database. These are typically related to full-length nucleic acid sequences that code for the production of a protein or peptide. One skilled in the art will recognize that identification of full-length sequences is not necessary from an analytical point of view. That is, portions of the sequences or ESTs can be selected according to well-known principles for which probes can be designed to assess gene expression for the corresponding gene.

Example 1- Sample Handling and LCM.

[0040] Twenty-seven fresh frozen tissue samples were collected from patients who had surgery for a colorectal tumor. Nineteen of the samples were colorectal malignancy specimens, and eight of the samples were of normal colon mucosa. The tissues were snap frozen in liquid nitrogen within 20-30 minutes of harvesting, and stored at -80C° thereafter. For laser capture, the samples were cut (6µm), and one section was mounted on a glass slide, and the second on film (P. A.L.M.), which had been fixed onto a glass slide (Micro Slides Colorfrost, VWR Scientific, Media, PA). The section

mounted on a glass slide was after fixed in cold acetone, and stained with Mayer's Haematoxylin (Sigma, St. Louis, MO). A pathologist analyzed the samples for diagnosis and grade. The clinical stage was estimated from the accompanying surgical pathology and clinical reports, using the Dukes classification. The section mounted on film was after fixed for five minutes in 100% ethanol, counter stained for 1 minute in eosin/100% ethanol (100µg of Eosin in 100ml of dehydrated ethanol), quickly soaked once in 100% ethanol to remove the free stain, and air dried for 10 minutes.

[0041] Two of the colorectal adenocarcinomas were of grade 1, 10 of grade 2, and 5 of grade 3. One of the malignant samples was a carcinoid tumor of the caecum, and one a metastatic melanoma lesion. Two of the adenocarcinoma samples represented the mucinous subtype, and one the signet cell subtype. The Dukes staging of the adenocarcinomas divided them as follows: Dukes A: 2, Dukes B: 5, Dukes C: 7, Dukes D: 3. Six of the adenocarcinomas had been irradiated preoperatively.

[0042] Before use in LCM, the membrane (LPC-MEMBRANE PEN FOIL 1.35µm No 8100, P.A.L.M. GmbH Mikrolaser Technologie, Bernried, Germany) and slides were pretreated to abolish RNases, and to enhance the attachment of the tissue sample onto the film. Briefly, the slides were washed in DEP H₂O, and the film was washed in RNase AWAY (Molecular Bioproducts, Inc., San Diego, CA) and rinsed in DEP H₂O. After attaching the film onto the glass slides, the slides were baked at +120°C for 8 hours, treated with TI-SAD (Diagnostic Products Corporation, Los Angeles, CA, 1:50 in DEP H₂O, filtered through cotton wool), and incubated at +37°C for 30 minutes. Immediately before use, a 10µl aliquot of RNase inhibitor solution (Rnasin Inhibitor 2500U=33U/µl N211A, Promega GmbH, Mannheim, Germany, 0.5µl in 400µl of freezing solution, containing 0.15 mol NaCl, 10 mmol Tris pH 8.0, 0.25 mmol dithiothreitol) was spread onto the film, where the tissue sample was to be mounted.

[0043] The tissue sections mounted on film were used for LCM. Approximately 2000 epithelial cells/sample were captured using the PALM Robot-Microbeam technology (P.A.L.M. Mikrolaser Technologie, Carl Zeiss, Inc., Thornwood, NY), coupled into Zeiss Axiovert 135 microscope (Carl Zeiss Jena GmbH, Jena, Germany). The surrounding stroma in the normal mucosa, and the occasional intervening stromal components in cancer samples, were included. The captured cells were put in tubes in 100% ethanol and preserved at -80°C.

Example 2- RNA Extraction and Amplification.

[0044] Zymo-Spin Column (Zymo Research, Orange, CA 92867) was used to extract total RNA from the LCM captured samples. About 2 ng of total RNA was resuspended in 10 ul of water and 2 rounds of the T7 RNA polymerase based amplification were performed to yield about 50 ug of amplified RNA.

Example 3- cDNA Microarray Hybridization and Quantitation.

[0045] A set of cDNA microarrays consisting of approximately 20,000 human cDNA clones was used to test the samples. About 30 plant genes were also printed on the microarrays as a control for non-specific hybridization. Cy3-labeled cDNA probes were synthesized from 5 ug of aRNA of the LCM captured cells. The probes were purified with Qiagen's Nucleotide Removal Columns and then hybridized to the microarrays for 14-16 hours. The slides were washed and air-dried before scanning. cDNA microarrays were scanned for cy3 fluorescence and ImaGene software (Biodiscovery, Los Angeles, CA) was used for quantitation. For each cDNA clone, four measurements were obtained using duplicate spots and duplicate arrays and the intensities were averaged.

[0046] cDNAs were printed on amino silane-coated slides (Coming) with a Generation III Micro-array Spotter (Molecular Dynamics). The cDNAs were PCR amplified, purified (Qiagen PCR purification kit), and mixed 1:1 with 10 M NaSCN printing buffer. Prior to hybridization micro-arrays were incubated in isopropanol at room temperature for 10 min. The probes were incubated at 95°C for 2 min, at room temperature for 5 min, and then applied to three replicate slides. Cover slips were sealed onto the slides with DPX (Fluka) and incubated at 42°C overnight. Slides were then washed at 55°C for 5min in 1X SSC/0.2% SDS and 0.1X SSC/0.2% SDS, dipped in 0.1X SSC and dried before being scanned by a GenIII Array Scanner (Molecular Dynamics). The fluorescence intensity for each spot was analyzed with AUTOGENE software (Biodiscovery, Los Angeles).

[0047] Chip intensities were linearly normalized forcing the intensity reading at the 75th percentile equivalent to a value of 100 on each chip. Every gene on the chip was normalized to itself by dividing the intensity reading for that gene by the median of the gene's expression value readings over all the samples. Prior to clustering, genes that did not have an intensity reading of 100 or greater in at least one sample were filtered out in order to limit the background affect on the similarity metrics. A set of 6,225 genes was selected for clustering analysis. Hierarchical clustering was performed using correlation as a measure of similarity, which groups together samples with genes that are showing positive changes at the same time without any consideration for negative changes (Silicon Genetics, Sunnyville, CA). Each of the major nodes in the dendrogram was then considered a subgroup of samples. Differentially expressed genes were identified by comparing each tumor subgroup to the normal group. The selection was based on a signal to noise measurement threshold given by

$$1 \leq \left| \frac{(\mu_t - \mu_n)}{(\sigma_t + \sigma_n)} \right|,$$

where μ_t is the mean of the tumor subset, μ_n is the mean of the subset of normal samples, and $\sigma_t + \sigma_n$ represent the combined standard deviations. The within-group coefficient of variation of the intensity readings of a gene had to be less than 0.33, for the gene to be included in the pair-wise comparisons. The median of the tumor group over the median of the normal group had to be greater than, or equal to 2 for up-regulation, and less than, or equal to 0.5 for down-regulation. If a gene met all the criteria, it was selected. The genes selected in all the comparisons were considered consistently dysregulated in colorectal cancer. The p-values for the statistical significance were calculated using a T-test assuming unequal variance. The gene set for clustering was also subjected to principal component analysis (PCA) using a software package (Partek, St Louis, MO). The data was then projected onto the reduced 3-dimensional space. The normal and tumor colorectal samples were represented by the projected expression levels.

[0048] A list of genes with large up-regulated differentials was created to distinguish between the tumor and normal samples. One-hundred and twenty-three genes were preselected by using

$$0.5 \leq \left| \frac{(\mu_t - MAX_n)}{(\sigma_t + \sigma_n)} \right|$$

as a signal to noise cutoff. A ratio equal to, or greater than 1.5 was the minimal criterion for up-regulation. Genes were also included if

$$0.9 \leq \left| \frac{(\mu_t - \mu_n)}{(\sigma_t + \sigma_n)} \right|.$$

A portfolio of four genes was established, each having at least a three fold expression differential between tumor and normal cells.

[0049] Differentially Expressed Genes in Colorectal Cancer. Thirty-nine genes were differentially expressed in all tumor samples as compared to normal colon mucosa. Thirty-seven of them were significantly down-regulated in all the tumors, except for an outlier. Two of them were up-regulated. The identities of the genes were verified by sequencing the cDNA clones placed on the microarray. Results are shown in Table 1.

Table 1

Modulated Genes					
ACCESSION	GENE DESCRIPTION	MEAN SIGNAL INTENSITY (NORMAL)	MEAN SIGNAL INTENSITY (TUMOR)	P-VALUE	
AF071569	CaM kinase II gene subtype delta 2.	93	39	4.64E-09	Seq. ID No. 1
AB014530	Homo sapiens mRNA for KIAA0630 protein	108	50	4.83E-07	Seq. ID No.2
AK000319	Human cDNA KIAA0630	236	69	7.84E-06	Seq. ID No.3
U81504	beta-3A-adaptin subunit of the AP-3 complex mRNA,	241	75	3.52E-05	Seq. ID No.4
AB011166	Human cDNA KIAA0594	116	55	3.53E-05	Seq. ID No.5
AB040914	Human cDNA KIAA1481	187	59	8.85E-05	Seq. ID No.6

Table 1 (continued)

Modulated Genes					
ACCESSION	GENE DESCRIPTION	MEAN SIGNAL INTENSITY (NORMAL)	MEAN SIGNAL INTENSITY (TUMOR)	P-VALUE	
AK025205	Human cDNA FLJ21552	322	97	0.00013	Seq. ID No.7
AJ278219	Fatty acid hydroxylase	143	53	0.00011	Seq. ID No.8
AB046854	Human cDNA KIAA1634	142	59	0.00020	Seq. ID No.9
R00585	Unknown	149	57	1.28E-09	Seq. ID No.10
S45844	Spi-B transcription factor	140	43	0.00043	Seq. ID No.11
X98311	Carcinoembryonic antigen family member 2 (CGM2)	6137	223	0.00044	Seq. ID No.12
BAA78050	NADPH oxidoreductase homolog	153	84	0.00048	Seq. ID No.40
N72128	Unknown	164	77	0.00068	Seq. ID No.13
AB040955	Human cDNA KIAA1552	334	120	0.00067	Seq. ID No.14
AF125101	HSPC040 protein	363	115	0.0011	Seq. ID No.15
AB023229	Human cDNA KIAA1012	263	88	0.00099	Seq. ID No.16
N95761	α -L-fucosidase gene	429	104	0.00047	Seq. ID No.17
AK025033	Human cDNA FLJ21380	180	85	0.0010	Seq. ID No.18
L10844	Human cellular growth regulating protein	206	101	0.0013	Seq. ID No.19
H96534	H.sapiens mRNA for gp25L2 protein.	147	58	0.0015	Seq. ID No.20
AK001521	Human cDNA FLJ10659	157	60	0.0019	Seq. ID No.21
AF151039	HSPC205 protein	117	60	0.0017	Seq. ID No.22
AF052059	SEL 1L protein	168	53	0.0016	Seq. ID No.23
N24597	Unknown	166	62	0.0016	Seq. ID No.24
AK001950	Inner centromere protein	148	64	0.0029	Seq. ID No.25
BAA02649	Macrophage scavenger receptor type I	118	44	0.0031	Seq. ID No.41
N75004	Unknown	98	48	0.0031	Seq. ID No.26

Table 1 (continued)

Modulated Genes					
ACCESSION	GENE DESCRIPTION	MEAN SIGNAL INTENSITY (NORMAL)	MEAN SIGNAL INTENSITY (TUMOR)	P-VALUE	
W16916	Human cDNA KIAA0260	162	61	0.0037	Seq. ID No.27
X52001	H.sapiens endothelin 3 mRNA.	89	33	0.0042	Seq. ID No.28
T50788	Unknown	364	102	0.0059	Seq. ID No.38
AJ005866	Putative Sqv-7 like protein	381	163	0.0049	Seq. ID No.29
AF113535	MAID protein	218	100	0.0053	Seq. ID No.39
AB037789	Human cDNA KIAA1368	164	62	0.0068	Seq. ID No.30
M33987	Carbonic anhydrase	652	46	0.0074	Seq. ID No.31
M77830	Desmoplakin 1 (DPI)	184	81	0.0092	Seq. ID No.32
H81220	EST domain transcription factor ELF1	113	55	0.017	Seq. ID No.33
AF000592	Human chromosome 21q11-q21 genomic clone	33	69	1.16E-05	Seq. ID No.35
AK021701	Human cDNA FLJ11639	31	63	0.00070	Seq. ID No.36

Example 4: Optimized Portfolio for Colorectal Tumors.

[0050] The mean-variance optimization algorithm was used to generate a multiple genebased signature, where the genes that are included can be used in combination to distinguish between the normal and tumor samples. Intensity measurements were processed using the samples and microarrays described in Examples 1-3. The data to be analyzed was first preselected based on a pre-specified 5-fold differential between tumor and normal cells. The expression data from genes preselected according to this criteria were then used as follows. The mean and standard deviation of the intensity measurements for each gene were calculated using the non-metastatic samples as the baseline. A discriminating value of $X \cdot (\text{Standard Deviation} + \text{Mean})$ was then calculated for each baseline gene (X was assigned a value of 3). This value was used to ensure the resulting portfolio would be stringent. A ratio of the discriminating value to the baseline value was then calculated for each metastatic sample. This ratio was then converted to a common logarithm. This data was then imported into Wagner Software which produced an efficient frontier from which a portfolio of 4 genes was selected. The set included an unknown sequence, procollagen type I, large subunit of ribosomal protein L21 and fibronectin. These genes are identified as Seq. ID No. 42, Seq. ID No. 43, Seq. ID No. 44 and Seq. ID No. 45. Alternatively, a combination of genes used to make up the portfolio can be used to produce diagnostic information that is useful for making clinical decisions regarding colorectal cancer. This is particularly beneficial in the case when a combination of genes selected from the portfolio are combined with additional markers (genetic or not).

Optimized Gene Portfolio:

[0051]

5 >gil1264443|gb|N92134.1|N92134 za23f09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE: 293417 5' similar to gb|IM87908|HUMALNE32 Human carcinoma cell-derived Alu RNA transcript, (rRNA); gb: X57025_mal INSULIN-LIKE GROWTH FACTOR IA PRECURSOR (HUMAN)
 >gil2221047|gb|AA490172.1|AA490172 ab06b08.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:839991 3' similar to gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN PRECURSOR (HUMAN)
 10 >gil2188918|gb|AA464034.1|AA464034 zx86b09.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:810617 5' similar to SW:RL21_HUMAN P46778 60S RIBOSOMAL PROTEIN L21.
 >gil834491|gb|I62612.1|I62612 yi12d01.s1 Soares placenta Nb2HP Homo sapiens cDNA clone IMAGE:139009 3' similar to gb:X02761_cds1 FIBRONECTIN PRECURSOR (HUMAN);

15 **[0052]** Using a different set of criteria but the same method, a further four gene portfolio was selected by the software. These are Seq. ID no. 46, Seq. ID No. 47, Seq. ID No. 48 and Seq. ID No. 49. Two genes overlap with the first four-gene portfolio. The two optimized portfolios can also be combined to form a six-gene portfolio.

Optimized Gene Portfolio:

[0053]

>gil2114953|gb|AA431245.1|AA431245 zw78d06.r1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE: 782315 5' similar to WP:F36H1.2 CE05814 ANKYRIN LIKE
 25 >gil2156172|gb|AA443497.1|AA443497 zw34d03.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone IMAGE:771173
 >gil2221047|gb|AA490172 ab06b08.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE: 839991 3' similar to gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN PRECURSOR (HUMAN)
 30 >gil1264443|gb|N92134.1|N92134 za23f09.r1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone IMAGE: 293417 5' similar to gb|IM87908|HUMALNE32 Human carcinoma cell-derived Alu RNA transcript, (rRNA); gb: X57025_mal INSULIN-LIKE GROWTH FACTOR IA PRECURSOR (HUMAN);

35

40

45

50

55

SEQUENCE LISTING

<110> WANG, YIXIN
 5 <120> COLORECTAL CANCER DIAGNOSTICS
 <130> CDS 267 US NP
 <140> TBD
 10 <141> 2003-03-21
 <150> 60/368,798
 <151> 2002-03-29
 <160> 49
 15 <170> PatentIn version 3.1
 <210> 1
 <211> 1500
 <212> DNA
 20 <213> human
 <400> 1
 atggcttcga ccaccacctg caccagggtc acggacgagt atcagctttt cgaggagctt
 60
 ggaaaggggg cattctcagt ggtgagaaga tgtatgaaaa ttcctactgg acaaggatat
 25 120
 gctgccaaaa ttatcaacac caaaaagctt tctgctaggg atcatcagaa actagaaaga
 180
 gaagctagaa tctgccgtct ttggaagcac cctaattattg tgcgacttca tgatagcata
 30 240
 tcagaagagg gcttttacta cttgggtgtt gatttagtta ctggagggtga actgtttgaa
 300
 gacatagtgg caagagaata ctacagtga gctgatgcc gtcattglat acagcagatt
 35 360
 ctagaaagtg ttaatcattg tcacctaaat ggcatagttc acagggacct gaagcctgag
 420
 aatttgcttt tagctagcaa atccaaggga gcagctgtga aattggcaga ctttggctta
 40 480
 gccatagaag ttcaagggga ccagcaggcg tggtttggtt ttgctggcac acctggatat
 540
 ctttctccag aagttttacg taaagatcct tatggaaagc cagtggatal gtgggcatgt
 45 600
 ggtgtcattc tctatattct acttgtgggg tatccacct tctgggatga agaccaaac
 660
 agactctatc agcagatcaa ggctggagct tatgatcttc catcaccaga atgggacacg
 50 720
 gtgactcclg aagccaaaga cctcatcaat aaaatgctta ctatcaacct tgccaaacgc
 780
 atcacagcct cagaggcact gaagcaccca tggatctgtc aacgttctac tgttgcttcc
 55 840

EP 1 355 149 A2

atgatgcaca gacaggagac tgtagactgc ttgaagaaat ttaatgctag aagaaaacta
900

5 aaggggtgcc a tcttgacaac tatgctggct acaaggaatt tctcagcagc caagagtttg
960

ttgaagaaac cagatggagt aaaggagtca actgagagtt caaatacaac aattgaggat
1020

10 gaagatgtga aagcacgaaa gcaagagatt atcaaagtca ctgaacaact gatcgaagct
1080

atcaacaatg gggactttga agcctacaca aaaatctgtg acccaggcct tactgctttt
1140

15 gaacctgaag ctttgggtaa tttagtggaa gggatggatt ttcaccgatt ctactttgaa
1200

aatgctttgt ccaaaagcaa taaaccaatc cacactatta ttctaaaccc tcatgtacat
1260

20 ctggtagggg atgatgccgc ctgcatagca tatattaggc tcacacagta catggatggc
1320

agtggaatgc caaagacaat gcagtcagaa gagactcgtg tgtggcaccg ccgggatgga
1380

25 aagtggcaga atgttcattt tcatcgctcg gggtcaccaa cagtaccat caagccaccc
1440

tgtattccaa atgggaaaga aaacttctca ggaggcacct ctttgtggca aaacatctga
1500

30 <210> 2
<211> 5761
<212> DNA
<213> human

35 <400> 2
cacaccgcag tatgcgggtgc cctttactct gagctgcgca gccggccggc cggcgctggg
60

tgaacagact gccgctglac tggcgtggcc tggagggact cagcaaattc tctgccttc
120

40 aacttggcaa cagttgcctg gggtagctct acacaactct gtccagccca cagcaatgat
180

tccagaggcc atggggagtg gacagcagct agctgactgg aggaatgcc actctcatgg
240

45 caaccagtac agcactatca tgcagcagcc atccttgetg actaaccatg tgacattggc
300

cactgctcag cctctgaatg ttggtgttgc ccatgttgtc agacaacaac aatccagttc
360

50 cctcccttcg aagaagaata agcagtcagc tccagtctct tccaagtcct ctctagatgt
420

tctgccttcc caagtctatt ctctggttgg gagcagtcce ctccgcacca catcttctta
480

55 taattccttg gtccctgtcc aagatcagca tcagcccatc atcattccag atactcccag
540

ccttctgtg agtgtcatca ctatccgaag tgacactgat gaggaagagg acaacaaata
 600
 5 caagcccagt agctctggac tgaagccaag gtctaattgc atcagttatg tcaactgtcaa
 660
 tgattctcca gactctgact cttctttgag cagcccttat tccactgata ccttgagtgc
 720
 10 tctccgaggc aatagtggat ccgttttgga ggggcctggc agagt.tgtgg cagatggcac
 780
 tggcaccgc actatcattg tgcctccact gaaaactcag cttggtgact gcactgtagc
 840
 15 aaccagggc tcaggtctcc tgagcaataa gactaagcca gtcgcttcag tgagtgggca
 900
 gtcatctgga tgctgtatca cccccacagg gtatcgagct caacgcgggg ggaccagtgc
 960
 20 agcacaacca ctcaatctta gccagaacca gcagtcatcg gcggctccaa cctcacagga
 1020
 gagaagcagc aaccagccc cccgcaggca gcaggcgttt gtggcccctc tctcccaagc
 1080
 25 cccctacacc ttccagcatg gcagcccgct acactcgaca gggcaccac accttgcccc
 1140
 ggcccctgct cacctgcaa gccaggetca tctgtatacg tatgtgccc cgacttctgc
 1200
 30 tgctgcactg ggctcaacca gctccattgc tcattctttc tccccacagg gttcctcaag
 1260
 gcattgtgca gcctatacca ctaccctag cactttgggtg caccaggctc ctgtcagtgt
 1320
 35 tgggcccage ctctcactt ctgccagcgt ggcccctgct cagtaccaac accagtttgc
 1380
 caccaatcc tacattgggt cttcccagg ctcaacaatt tacactggat acccgctgag
 1440
 40 tctaccaag atcagccagt attcctactt atagttgggtg agcatgaggg aggaggaatc
 1500
 atggctacct tctcctggcc ctgcgttctt aatattgggc tatggagaga tctccttta
 1560
 45 cctcttgaa atttcttagc cagcaacttg ttctgcaggg gccactgaa gcaqaagggt
 1620
 tttctctggg ggaacctgtc tcagtgttga ctgcattgtt gtagtcttcc caaagtttgc
 1680
 50 cctattttta aattcattat tttt.tgaca gtaatttttg tacttggaag agttcagatg
 1740
 cccatcttct gcagttacca aggaagagag attgttctga agttaccctc tgaaaaatat
 1800
 55 tttgtctctc tgacttgatt tctataaatg cttttaaaaa caagtgaagc cctctttat
 1860

EP 1 355 149 A2

ttcattttgt gttattgtga ttgctggtca ggaaaaatgc tgatagaagy agllgaaatc
 1920
 5 tgatgacaaa aaaagaaaaa ttactttttg tttgtttata aactcagact tgcctatttt
 1980
 attttaaaaag cggcttacac aatctccctt ttgtttattg gacattttaa cttacagagt
 2040
 10 ttcagttttg ttttaaatgac atattatact taatgggcaa ttgttatttt tgcaaaaactg
 2100
 gttacgtatt actctgtgtt actattgaga ttctctcaat tgctcctgtg tttgttataa
 2160
 15 agtagtgttt aaaaggcagc tcaccatttg ctggttaactt aatgtgagag aatccatata
 2220
 tgcgtgaaaa caccaagtat tcttttttaa tgaagcacca tgaattcttt tttaaattat
 2280
 20 tttttaaaag tctttctctc tctgattcag cttaaatttt tttatcgaaa aagccattaa
 2340
 ggtgggttatt attacatggt ggtgggtggt ttattatatg caaaatctct gtctattatg
 2400
 25 agatactggc attgatgagc tttgcctaaa gatlagtatg aattttcagt aatacacctc
 2460
 tgttttgctc atctctccct tctgttttat gtgatttggt tggggagaaa gctaaaaaaa
 2520
 30 cctgaaacca gataagaaca tttcttgtgt atagctttta tacttcaaag tagcttcctt
 2580
 tgtatgccag cagcaaattg aatgctctct tattaagact tatataataa gtgcatgtag
 2640
 35 gaattgcaaa aaatatttta aaaatttatt actgaattta aaaatatttt agaagttttg
 2700
 taatgggtgt gttttaatat ttacataat taaatatgta catattgatt agaaaaatat
 2760
 40 aacaagcaat ttttcctgct aacccaaaat gttatttgta atcaaatgtg tagtgattac
 2820
 acttgaattg tgtacttagt gtgtatgtga tctccagtg ttatcccgga gatggattga
 2880
 45 tgtctccatt gtatttaaac caaaatgaac tgatacttgt tggaatgtat gtgaactaat
 2940
 tgcaattata ttagagcata ttactgtagt gctgaatgag caggggcatt gcctgcaagg
 3000
 50 agaggagacc cttggaattg ttttgcacag gtgtgtcttg tgaggagttt ttcagtgtgt
 3060
 gtctcttctt tccctttctt cctccttccc ttattgtagt gccttatatg ataattgtat
 3120
 55 ggttaataga gtttacagtg agcttgccctt aggatggacc agcaagcccc cgtggaccct
 3180

EP 1 355 149 A2

aagttgttca ccgggattta tcagaacagg attagtagct gtatttgtta atgcattgtt
3240

5 ctcagtttcc ctgccaacat tgaaaaataa aaacagcagc ttttctcctt taccaccacc
3300

tctaccctt tccattttgg attctcggcl gagttctcac agaagcattt tccccatgtg
3360

10 gctctctcac tgtgcgttgc taccttgctt ctgtgagaat tcaggaagca ggtgagagga
3420

gtcaagccaa tattaaatat gcattctttt aaagtatgtg caatcacttt tagaatgaat
3480

15 ttttttttcc ttttcccatg tggcagtcct tcctgcacat agttgacatt cctagtaaaa
3540

tatttgcttg ttgaaaaaaaa catgttaaca gatgtgttta taccaaagag cctgttlat
3600

20 tgcttaccat gtccccatac tatgaggaga agttttgtgg tgccgctggg gacaaggaac
3660

tcacagaaag gtttcttagc tgggtgaagaa tatagagaag gaaccaaagc ctgttgagtc
3720

25 attgaggctt ttgaggtttc ttttttaaca gcttgtatag tcttggggcc cttcaagctg
3780

tgaaattgtc ctgtactct cagctcctgc atggatctgg gtcaagtaga aggtactggg
3840

30 gatggggaca ttctgcca taaaggattt ggggaaagaa gattaatcct aaaatacagg
3900

tgtgttccat ccgaattgaa aatgatatat ttgagatata attttaggac tggttctgtg
3960

35 tagatagaga tgggtgtcaag gaggtgcagg atggagatgg gagatttcat ggagcctggg
4020

cagccagctc tgtaccaggt tgaacaccga ggagctgtca aagtatttgg agtttcttca
4080

40 ttgtaaggag taagggttc caagatgggg caggtagtcc gtacagccta ccaggaacat
4140

gttgtgtttt ctttattttt taaaatcatt atattgagtt gtgttttcag cactatattg
4200

45 gtcaagatag ccaagcagtt tgtataattt ctgtcactag tgtcatacag ttttctggtc
4260

aacatgtgtg atcttttgtt ctcctttttg ccaagcacat tctgaltttc ttgttggaa
4320

50 acaggtctag tttctaaagg acaaattttt tgttccttgt cttttttctg taagggacaa
4380

gatttgttgt ttttgaaga aatgagatgc aggaagaaa accaaatccc attcctgcac
4440

55 ccagtcctaa taagcagata ccacttaaga taggaqtcta aactccacag aaaaggataa
4500

EP 1 355 149 A2

5 taccaagagc ttgtattggt accttagtca cttgcctagc agtgtgtggc ttt.aaaaact
 4560
 agagattttt cagtcttagt ctgcaaactg gcatttccga ttttccagca taaaaatcca
 4620
 cctgtgtctg ctgaatgtgt atgtatgtgc tcaactgtggc tttagattct gtccttggg
 4680
 10 ttagccctgt tggccctgac aggaaggag gaagcctggt gaatttagtg agcagctygc
 4740
 ctgggtcaca gtgacctgac ctcaaaccag ctttaaggctt taagtcctct ctcagaactt
 4800
 15 ggcatttcca acttcttctt ttcgggtga gagaagaagc ggagaagggt tcagtgtagc
 4860
 cactctgggc tcataggagc acttggtcac tccagagttt ttaatagctc ccaggaggtg
 4920
 20 atattatttt cagtgtcag ctgaaatacc aaccacagga ataagaactc catttcaaac
 4980
 agttctggcc attctgagcc tgcttttgtg attgtctatc cattgtctct cactagaggg
 5040
 25 gctaagcttg actgccctta gccaggcaag cacagtaatg tgtgttttgt tcagcattat
 5100
 tatgcaaaaa ttcactagtt gagatggtt gttttaggat aggaaatgaa attgcctctc
 5160
 30 agtgacagga gtggcccgag cctgcttctt attttgattt tttttttttt taactgatag
 5220
 atggtgcagc atgtctacat gggtgtttgt tgctaaactt tatataatgt gtggtttcaa
 5280
 35 ttcagcttga aaaataatct cactacatgt agcagtacat tatatgtaca ttatatgtaa
 5340
 tgttagtatt tctgctttga atccttgata ttgcaatgga attcctactt tattaaatgt
 5400
 40 atttgatatg ctagttattg tgtgcgattt aaactttttt tgctttctct ctttttttgg
 5460
 ttgtgcgctt tcttttataa caagcctcta gaaacagata gtttctgaga attactgagc
 5520
 45 tatgtttgta atgcagatgt acttagggag tatgtaaaat aatcatttta acaaagaaa
 5580
 tagatattta aaatttaata ctaactatgg gaaaagggtc cattgtgtaa aacatagttt
 5640
 50 atctttggat tcaatgtttg tctttgggtt tacaaagtag cttgtatttt cagtattttc
 5700
 tacataatat ggtaaaatgt agagcaattg caatgcatca ataaaatggg taaattttct
 5760
 55 g
 761

5

5 <210> 3
 <211> 2129
 <212> DNA
 <213> human

10 <400> 3
 ctgtattgag acaaaggaag ggatctgtca gaaagcaaca cttgttatct tgggcttggc
 60
 agcaaggaag aggacaggta gtggagatcc tgcaatctga aaagcagact gaaagggtgac
 120
 aaagaagctg aagatgggtg gtggagagag gtataacatt ccagcccctc aatctagaaa
 180
 15 tgttagtaag aaccaacaac agcttaacag acagaagacc aaggaacaga attcccagat
 240
 gaagattgtt cataagaaaa aagaaagagg acatggttat aatcatcag cagctgcctg
 300
 20 gcaggccatg caaaatgggg ggaagaacaa aaattttcca aataatcaaa gttggaattc
 360
 tagcttatca ggtcccaggt tacttttttaa atctcaagct aatcagaact atgctggtgc
 420
 25 caaatttagt gagccgcat caccaagtgt tcttccaaa ccaccaagcc actgggtccc
 480
 tgtttccttt aatccttcag ataaggaaat aatgacattt caacttaaaa cttacttaa
 540
 30 agtacaggta taaaataaga caaatgttta aatttagtta tgttcacggg tagttgtcaa
 600
 ttggtctgaa acaaatttgc tagggaatct atttgtgtag aactaattaa tgtaaaaaaa
 660
 35 atagaccatc tcgtgttggtg tgcactgtga tataatggta gtatcagtc aacttaaaact
 720
 aatgattgta cttgatatta agtgttctca actgagtaac ttttaagtgg aaaccaagtt
 780
 40 tagatttggg gagtggtaaa ggaatcagct ttttctattg ttaggggaag acagtaattt
 840
 atcattcatg gaccagtaga ttgttgaaag ttggtgaatc ggattataag cttctagcta
 900
 45 acacaaggat tcagaattag gtaaacatct gaaggtttag tatattagaa acacccaaac
 960
 cagtaatatg ctaacctgat gcactgctga aagaaaatgt gaatttttcg taataattgc
 1020
 50 attttagtga attgtacagt ggggtgaaa ggcatttggg gctcattaga atgagacata
 1080
 gtacacccca atggccctgt ttattaaatg tagtggatta agtgtctgtc aacaaataca
 1140
 55 ccaaaacat tttttataga aacagtattt aatggtcact caatagcttt caaaatacat

EP 1 355 149 A2

1200

ttttgtatta cagcactgca caagctattc taatagtgc ctggcctcat cattcctgca
1260

5

aagccttgctt tggggagttg galaatgtga aaattttaag tacctagggg agaaagagcc
1320

atgtaaatat ctgtaataaa cttgtagcat atgtaaagtt ttcttggcct ttatcttaca
1380

10

aaaatggagt attttagtat gaatttgctg aatgtaagac cgtggactgt tttttataat
1440

atggcctaata tttaaaggta caaaataact tgtttttaaa gtttgcctt gtgctaaagt
1500

15

gccagtgtat gtatgttata cttgatttgg ttgtaaacta tatttcaaag taaaccctag
1560

tgtaataagt ttataacta aaaagggtta agctgctaaa actattttta agagatgtga
1620

20

aatgcagtat gggactatct ttttttctc ctctaagccc aaagattaac tagagtcctt
1680

ccaaccttat agattgttgg ctttcacaat cttataacct aggatacagg lagtttcgag
1740

25

tatggtgcca gtgatgtttt gtttttgttt ggtcaagggg taggtgcaac ccaatggacc
1800

acttatgcaa aagatgtaaa ctcttgcata atacattgat aacatgtttt gccaaactta
1860

30

aatgcttaaa cataagcgaa accagtagca agtatgtggg tcagcttaaa aattttgatt
1920

gttaatgccc tattttctaa ttggcacct cttgatgcct aagcaggtaa gcagatgcct
1980

35

aagctgtatt tctccaaata aatcaagatg aagtactgcc caagttaaatt attgatagcc
2040

taaagacaag ttatgtagt acttaatgta catgatatga agcataaaat taaataaaat
2100

40

ttttcccat tgaaaaaaaa aaaaaaaaaa
2129

<210> 4

45

<211> 3950

<212> DNA

<213> human

<400> 4

cgagaactag ttttgttccg tgcctcttgg actggaacct ttgggagaga acccccggca
60

50

ggaccaaccc cgcacccgcc agcacccggg caatgtccag caatagtttt ccttacaatg
120

agcagtcagg aggaggggag gcgacggagc tgggtcagga ggcgacctca accatttccc
180

55

EP 1 355 149 A2

cctcgggggc cttcggcctc tttagcagcg atttgaagaa gaatgaagat ctaaagcaaa
240

5 tgltagagag caacaaagat tctgctaaac tggatgctat gaagcggatt gttgggatga
300

ttgcaaaaagg gaaaaatgca tctgaactgt ttcctgctgt tgtgaagaat gtggccagta
360

10 aaaatattga gatcaagaag ttggtatatg tttacctggg tcgatatgct gaagaacagc
420

aggatcttgc actcctgtcc ataagcactt ttcagcgagc tctgaaggac ccaaaccaac
480

15 taattcgtgc aagcgctttg agagttctgt caagtattag agtgccaatt attgtacct
540

tcatgatgct tgctattaag gaagcttctg ctgacttacc accatatgtt aggaagaatg
600

20 cagcccatgc aatacaaaaa ttatacagcc ttgatccaga gcagaaggaa atgttaattg
660

aagtaattga aaaacttctg aaagataaaa gcacattggg agctggcagt gttgtgatgg
720

25 cttttgaaga agtatgcccg gacagaatag atctgattca taaaaattac cgcaagctat
780

gtaacttact agtggatggt gaagagtggg ggcaggttgt cataatccac atgctaactc
840

30 gatatgctcg gacacagttt gtcagccctt ggaaagaggg tgatgaatta gaagacaatg
900

gaaagaattt ctacgaatct gatgatgac agaaggaaaa gactgacaaa aagaagaagc
960

35 cgtatactat ggatccagat catagactct taattagaaa tacaagcct ttgcttcaga
1020

gcaggaatgc tgcggtgggt atggcagttg ctcagctgta ttggcacata tcacccaaat
1080

40 ctgaagctgg cataatttct aaatcactag tgcgtttact tcgtagcaat agggaggtgc
1140

agtatattgt cctacaaaat atagcaacta tgtcaattca aagaaagggg atgtttgaac
1200

45 cttatctgaa gagtttctat gttaggtcaa ctgatccaac tatgatcaag aactgaagc
1260

ttgaaatttt gacaaacttg gcaaatgaag ccaacatac aactcttctt cgagaatttc
1320

50 agacctatgt gaaaagccag gataaacaat ttgcagcagc cactattcag actataggca
1380

galglgcaac caacatcttg gaagtcactg acacgtgcct caatggcttg gtctgtctgc
1440

55 tgtccaacag ggatgaaata gttgttgctg aaagtgtggg tgttataaag aaattactgc
1500

EP 1 355 149 A2

aaatgcaacc tgcacaacat gylgaaatta llaaacatat ggccaaactc ctggacagta
 1560
 5 tcaactgttcc tgttgctaga gcaagtattc tttggctaata tggagaaaac tgtgaacqag
 1620
 ttcctaaaaat tgccctgat gttttgagga agatggctaa aagcttcact agtgaagatg
 1680
 10 atctggtaaa actgcagata ttaaactctgg gagcaaaatt gtaatttaacc aactccaaac
 1740
 agacaaaatt gcttaccag tacatattaa atctcgcaa gtatgatcaa aactacgaca
 1800
 15 tcagagaccg tacaagattt attaggcagc ttattgttcc gaatgaaaag agtggagctt
 1860
 taagtaaata tgccaaaaaa atattcctag caaaaagcc tgcaccactg cttgagtctc
 1920
 20 cttttaaaga tagagatcat ttccagcttg gcaccttctc tcatactctc aacattaaag
 1980
 ctactgggta cctggaatta tctaattggc cagaggtggc gcccgacca tcagtccgaa
 2040
 25 atgtagaagt aatagagttg gcaaaagaat ggacccagc aggaaaagca aagcaagaga
 2100
 attctgctaa gaagttttat tctgaatctg aggaagagga ggactcttct gatagtagca
 2160
 30 gtgacagtga gagtgaatct ggaagtgaat gtggagaaca aggcgaaagt ggggaggaag
 2220
 gagacagcaa tgaggacagc agtgaggact cctccagtga gcaggacagt gagagtggac
 2280
 35 gggagtcagg cctagaaaac aaaagaacag ccaagaggaa ctcaaaagcc aaaggaaaaa
 2340
 gtgattctga agatggggag aaggaaaatg aaaaatctaa aacttcagat tcttcaaag
 2400
 40 acgaatctag ttcaatagaa gacagttctt ccgattctga atcagagtca gaacctgaaa
 2460
 gtgaatctga atccagaaga gtcactaagg agaaagaaaa gaaaacaaag caagatagaa
 2520
 45 ctctcttac caaagatgtt tcaattctag atctggatga ttttaacca gtatccactc
 2580
 cagttgcact tcccacacca gctctttctc caagtttgat ggctgatctt gaaggtttac
 2640
 50 acttgtaac ttctcttca gtcactcagtg tcagtactcc tgcatttgta ccaacgaaaa
 2700
 ctacgtgct gcttcacga atgagtggaa aaggactagc tgccattat ttctttccaa
 2760
 55 gacagccttg catTTTTTggt gataagatgg tctctataca aataacactg aataacacta
 2820

EP 1 355 149 A2

ctgatcgaaa gatagaaaat atccacatag gggaaaaaaa aclicctata ggcattgaaa
2880

5 tgcattgtttt taatccaata gactctcttg agcctgaggg atccattaca gtttcaatgg
2940

gtattgactt ttgtgattct actcagactg ccagtttcca gttgtgtacc aaggatgatt
3000

10 gcttcaatgt taatattcag ccacctgttg gagaactgct tttacctgtg gccatgtcag
3060

agaaagattt taagaaagag caaggagtgc taacaggaat gaatgaaact tctgctgtaa
3120

15 tcattgctgc accacagaat ttcactccct ctgtgatctt tcagaagggt gtaaattgag
3180

ccaatglagg tgcagtccct tctggccagg ataatatata caggtttgca gctaaaactg
3240

20 tgcacagtgg gtcaltgatg ctagtccacag tggaaactgaa ggaaggctct acagcccagc
3300

ttatcataaa cactgagaaa actgtgattg gctctgttct gctgcgggaa ctgaagcctg
3360

25 tcctgtctca ggggtaacct gcttacatct ggactttaga atctggcaca caacaaaagt
3420

gcctggcatc cactactgct gcctttcatt tataataata gcccttccat ctggcagtgg
3480

30 gggtagaata cactcttgac attcttgtct cctgctttag aatgctagtg tgtatctatc
3540

atgtatgcaa tactttcccc ctttttgctt tgctaaccga agagcatata ttttactgtc
3600

35 agttgtctca actcttgaat ccatgtggcg ttttctctgt cctgctgctt cttttggcct
3660

cctcgttttc cttctctttt tcgacaatgg tagacatgaa tgagatattt aaagttcatt
3720

40 ggaaatcttc ttccctacag cagtaagcaa aaattagcaa agagatagtc taaatggcct
3780

ctcagcttgg tatgtgaaaa tgagatcaca tactttttta atccaaatac aaaagcatag
3840

45 tctctgcaag attttgttct ttgaatttct tgatattgta attgattatt gataactgtc
3900

atcatgaaat tatctctcaa taataagata aataaactag catatgaatc
3950

50 <210> 5
<211> 5191
<212> DNA
<213> human

<400> 5
55 gagaaagaaa aacagctcga gacctcatgc aaagagaaaa ctgagtatct acagaaaatg
60

gttcagagga atgaaagata taaacaagat gtggagaggt tctatgaacg gaagcgacat
120

5 ttagatttaa ttgagatgct tgaagcaaaa aggccatggg tggaatatga aaatgttcgt
180

caggaatatg aagaagtaaa actagtctgt gaccgagtga aggaagaggt cagaaaactt
240

10 aaagaagggc agattcctat aacatgtcga attgaagaaa tggaaaacga gcgtcacaat
300

ttggaggctc gaatcaaaga aaaggcaaca gatattaagg aggcattctca aaaatgcaaa
360

15 cagaagcaag atgttataga aaggaaagat aaacatattg aggaacttca gcaggcttta
420

atagtaaagc aaaatgaaga gcttgaccga cagaggagaa taggtaatac ccgcaaatg
480

20 atagaggatt tgcaaaatga actaaagacc acggaaaact gcgagaatct tcagccccag
540

attgatgcca ttacaaatga tctgagacgg attcaggatg aaaaggcatt atgtgaaggc
600

25 gaaataattg ataagcgaag agagagggaa actctagaga aggagaaaaa gagtgtggac
660

gatcatattg tacgttttga caatcttatg aatcagaagg aagataagct aagacagaga
720

30 ttccgtgaca cgtatgatgc tgttttatgg ctaagaaata acagagacaa atttaaacia
780

agagtctgtg agcccataat gctcacgac aatatgaaag ataataaaaa tgccaaatat
840

35 attgaaaatc atattccatc aaatgactta agagcctttg tatttgaaag tcaagaagat
900

atggagggtt tcctcaaaga ggttcgtgac aataaaaaat taagagtaaa tgctgttatt
960

40 gctcccaaga gtcatatgc agacaaagca cttcaagat ctttgaatga acttaaacia
1020

tacggatttt tctcttattt gagagaatta tttgatgcac ctgacctgt aatgagttac
1080

45 ctttgctgtc agtatcatat tcatgaagtt cctgtaggaa ctgaaaagac cagagaaaga
1140

attgaacggg taatacaaga aaccgatta aaacagattt atacagcaga agaaaagtat
1200

50 gtggtgaaaa ctctctttta ttcaaaciaa gttatttcta gtaacacatc tctaaaagta
1260

gcgcagtttc tcactgtcac tgtggaccta gagcagagaa gacacttaga agaacagcta
1320

55 aaggaaattc atagaaaatt gcaagcagtg gattcagggt tgattgcctt acgtgaaaca
1380

EP 1 355 149 A2

agcaaacatc tggagcacia agacaatgaa cttagacaaa agaagaagga gcttcttgag
1440

5 agaaaaacca agaaaagaca actggaacaa aaaatcagtt ccaaaactagg aagttttaaag
1500

ctgatggaac aggatacttg caatcttgaa gaggaagagc gaaaagcaag taccaaaatc
1560

10 aaagaaataa atgttcaaaa agcgaaactt gttaccgaat taacaaacct aataaaqatt
1620

tgtacttctt tgcataatac aaaagtagat ttaattctcc aaaatactac agtgatctct
1680

15 gagaagaaca aattagaatc agattalatg gccgcactct cacaactccg tcttacagag
1740

caacatttca ttgaattgga tgaaaataga cagagattat tgcagaaatg caaggaactt
1800

20 atgaaaagag ctaggcaagt atgtaacctg ggtgcagagc agactcttcc tcaagaatac
1860

cagacacaag taccacccat tccaaatgga cacaactccl cactcccat gggtttccaa
1920

25 gaccttccaa acacaltgga tgaaattgat gctttattaa ctgaagaaag atcaagagct
1980

tcctgcttca cgggactgaa tcttacaatt gttcaggaat atacaaaaag agaagaagaa
2040

30 atagaacagt taactgagga actaaaggga aagaaagtgt aactagatca atacagggaa
2100

aacatttcac aggtaaaaga aagggtggctt aatcctltta aagagctggt agaaaaaatt
2160

35 aatgaaaaat tcagcaattt ttttagttcc atgcagtgtg ctggtgaagt tgatctccat
2220

acagaaaatg aggaagatta tgataaatat ggaattcgaa ttagagtcaa atttcgaagt
2280

40 agtactcaac tgcataaatt aactcctcat catcaaagtg gaggtgaaag aagtgtttct
2340

accatgttat acttgatggc acttcaggag ctaaatagat gtccattcag agtagttgat
2400

45 gaaatcaatc agggaatgga cccaatcaat gaacggagag tgtttgaaat ggttgtaaat
2460

actgcctgta aagaaaatac atctcaatac ttttccataa caccaaagct cctgcaaaat
2520

50 ctctcttatt ctgaaaagat gacagttttg tttgtctaca atggccctca tatgctggaa
2580

ccaaacacat ggaatttaaa ggctttccaa aggcggcggc gccgtattac attcactcaa
2640

55 ccttcttaat aaaagtaaag agaggggaact tgggaatttt ttttgttaaa ttctgtttat
2700

EP 1 355 149 A2

aagtatggct caactgaata aaaggagatt cactaaaacg aaaagcagtt atttttggaa
2760

5 acctgctttt aaatacaaat aggttgataa tggaaactat aatgacctt ccaaaatagc
2820

agctggtagt aaaagttaag tcttcttcag tcttggttga acttgagttc ttggcactct
2880

10 gaccatgagt cattcagttc tcatgttaaa atgtacttaa tattacaaat caaaggtaga
2940

gtggaagaag ggtaatcac aagaagttac ttatatggtg gccctgagct ttaattgcag
3000

15 agtaacttta attactttta gagcctaaag atgactctag agcctaagtc ctagtctctc
3060

ccaatgttat atttaatttt aaaaaattga tatgaaaatg tctaattgat agtaataatt
3120

20 tatgacagat ctagtcatct cttcctatta aaaaagatta ccttatctcc agtaggaaat
3180

ggaattttat gggcctttta aagaaagttt tatgaaactt gatgctataa ttttattggt
3240

25 atttcaaggg gaaaaaagca ctggggttca aaaatggtag cagaactgct ttgaaatgct
3300

gcaaggtggc cactagatga tgcaaaatac aaccaaaga ttgactgaga ataaaattag
3360

30 gtgacaaggg tttttaaaga ataacctttt aaagtgtggg ggcaggggtt gctttttttt
3420

attttattta aagtcattta tttttacat cttacatttc taaaagcatt ttataattat
3480

35 ttttagtaag atttttctta aaatttcata tactggtttc tacaatttat atttgaaatt
3540

tctcagtgtt atgtaaagag tgatggaaaa gcattgattt ctttaaaacc gtaatgtttt
3600

40 tagaacttaa gcctataggg cttttcttac aatgttgatg taccattat cttagaaat
3660

ctagttaaaa ctgttttctt tcaccgcaa agaattaaat gggaaaatca tttgtttatc
3720

45 tctaagttat actaattagt agaaccaaac aaattatctt cttttaaaaa ataaatctta
3780

taggaaaaata gacagtccaa agtcatgtct ttgaacagtg gattggatct gtgccagtaa
3840

50 tgacaaaatt atttttttga cttgcttgcc tgaataaatt gaagaattgc tttcagtttg
3900

ggttttgtat attcttaagt agccattgaa atttatattc ttaactaggt caaaaaataa
3960

55 tgagccataa gtttatgtcc tctcacttag acattttctc tttaaaaagg tattttcttc
4020

EP 1 355 149 A2

tttataaaca ttttaaaaga gccttcctt cttaaaactaa ctccagtga tgaagtgtga
 4080
 5 aaatatttta aaatgacatt tttaactaata tgagcaagtc atgtaaacat tgaagaactt
 4140
 ggtaacatat tagtaaatgg atattaccaa atgttttcat cgtaattac ttgcgttcc
 4200
 10 accaaaatat ctttactaaa atgtgcttgg tgtagtttgt ttattgtcta aattagtlacc
 4260
 aglcatctta tttctgcaaa atgagtatca atgtgaaaaa gacacgtgaa gattaagcat
 4320
 15 gtttgaaaat aaaatggta attacatttc aatttacata ggccaacaac tgttccatac
 4380
 tttgtttgta aacatttaatt ttctctactg gacaaaatta atatttggct ttacattgaa
 4440
 20 ttttgagctg tgaagaataa attatgtatc attttagcat attaaacagt agtaagtcta
 4500
 gcacatagtc tcagccactt aaaacaaaag tttttttggt tgtttgtttg tttgtttttt
 4560
 25 tgagatggag tctcactctg ttgccaggc tggagtgcag tggcgtgatc tcggcttact
 4620
 gcaacctccg cctcccggt tcaagcgatt ctctgcctc agcctccca gtaactggga
 4680
 30 caacaggcgc gtcccaccac acccagctaa ttttttatac ttttagtaga gatggggttt
 4740
 cagcatattg gccaggctgg tctcgaaactc ctgacctgt gatccaccg cctcggcctc
 4800
 35 ccaaagtgtc gggattatag gcgtgagccc ctgcaccg ccaaaagttg atttttaatt
 4860
 acataaaaaat cgtaaaaact tctagtaaaa acttgatttg gtgaatacag ttatatatta
 4920
 40 aaaccttaag gtgacaagca ttttctatgc ctaaatcttc attgggtttgc ctggaaagag
 4980
 tctctgttaa aagattttcc atattcaaag taaaaggaaa gatttcttgc ttcctaattg
 5040
 45 tcttttgac acatgcctat tttctttgag gtataaacct ttagatgtga aaaatgtaat
 5100
 ttcattctgc tattgtgtgt gcttgtgtgt gtgtaattga aaaaactggg aaatcctgct
 5160
 50 ttgttggtta taaatcaata tttttatatt c
 5191
 <210> 6
 <211> 4755
 <212> DNA
 <213> human

<400> 6
 aagagatctt ccaggctctc agagccctgg gagggcgatt tccaggaaga ccacaatgcc
 60
 5
 aacctctgga ggaggctgga gagagaagyc claygccaga gcctgtcagg caactttggc
 120
 aagaccaagt cagccttctc atctctccag aacattcctg agagtctgag aagacacagc
 180
 10
 agcctggagc taggccgggg aacccaggag ggttaccctg ggggcaggcc cacctgtgca
 240
 gtcaacacca aggcagaaga ccctgggagg aaagccgctc ctgacctcgg gagccatctg
 300
 15
 gaccggcagg ttctctaccc gcggcccgag gggaggaccg gtgcctcggc ttctttcaac
 360
 agcacagacc caagtcccg aagagccgct gccccctcgc acccgcacac atccagtctg
 420
 20
 ggccggaggg ggcccgccc aggcagcgcc tcggctcttc agggctttca gtacggyaag
 480
 cccactgct cgggtctgga gaaggtctcc aaatlccagc agcgagagca agggagccag
 540
 25
 agaccgagtg tggycggctc tggttttggc cataactata ggccccacag gaccgtctca
 600
 acttcagta cttctgggaa tgacttcgag gagacaaaag cacacattcg tttctctgag
 660
 30
 tcagctgaac ccctaggcaa cggggagcag cacttcaaaa acggggagct gaagttggaa
 720
 gaggttccc ggcagccctg cyglcagcag ctgagcggag gagcgtcggg cagcggccgt
 780
 35
 ggccccaga ggccggacgc tcggctctc cgtagccaga gcaccttcca gctctccagc
 840
 gagccagaga gggagcccga gtggcgggac aggcccggt cggccgaatc gccctgctg
 900
 40
 gatgccccct tcagccgcgc ctaccggaac agcatcaagg acgcacagtc ccgtgtcttg
 960
 ggggccacct cctttcgacg tcgagacctg gagctggggg cggccgtggc gtcgaggtcc
 1020
 45
 tggcgccac ggcttctc ggccacgtg gggctgcgga gccccgagc gtcggcctcc
 1080
 gcctcccg c acagccccg ggagcggcac agcgtgacct ctgctgagg cgacctggcc
 1140
 50
 aggcccgctg cccctgccgc ccggagaggt gctcgccggc gcctgactcc cgagcagaag
 1200
 aagcgctcct actcggagcc cgagaagatg aacgaggtgg ggalcglgga ggaggccgaa
 1260
 55
 ccggcacccc tgggcccga gagaaatgg atgcgtttcc cggagagcag cgtggccgac

EP 1 355 149 A2

1320

5 cggcgccgtc tcttcgagcg cgatggcaag gcctgctcca cgctcagcct gtcggyggccc
1380

gagctgaagc agttccagca gagcgccctg gcggaclaca tccagcgcaa gaccggcaag
1440

10 cggcctacct ccgcccgcgg ctgcagcctc caggagcccg ggccactgcy tgagcgcgcc
1500

cagagtgcct acctccagcc cggccccgcy gcgctcgaag gctccggcct cgcctcggcc
1560

15 tccagcttga gctcactgcy ggagcccagc ctgcagcccc gcagggaggc cacgctcctg
1620

ccggccacag ttgcagaaac ccagcaggct ccccgagatc gcagcagctc cttcgccggt
1680

20 ggccgcccgc tccgggaacg gcgagcgggg gacctgctta gcggagcaaa cggtggaaca
1740

aggggcaccc agagagggga tgagaccccc agggagccat cctcctgggg ggccaggggc
1800

25 gggaagtcca tgtcggccga ggacctgctg gaacgctcgg acgtccttgc gggccctgtc
1860

catgtgaggt ccaggtcacc tcccggcacc gcagacaagc gccaggatgt gcttttgggg
1920

30 caagacagtg gctttggtct tgtgaaggat ccatgttatt tggttggtcc tggatctagg
1980

tcactcagtt gttcagaaag agggcaagaa gagatgctgc tgctcttcca ccatctcacc
2040

35 cctcgttggg gtggttcagg ctgcaaagcc attggtgatt cctccgttcc tagtgaatgt
2100

cctggaaccc tggaccatca gaggcaagcc agtaggacac cctgccccag gccaccactg
2160

40 gcaggaacgc aagggtcgtt cacagacacc agggctgcac cctgacccc aattggcacc
2220

cctctgcctt cagccattcc ctctggctac tgctcacagg acggtcagac agggcgacag
2280

45 cctctcccgc cctacacccc tgccatgatg cacagaagca atggtcacac cctgacccag
2340

cctcccggtc caagaggctg tgagggcgat ggcccagagc atggggtaga agagggaaag
2400

50 aggaagaggg tctcgctgcc tcagtggcca cctccttctc gagcaaagtg ggcccacgca
2460

gccagagagg acagccttcc tgaggaatcc tcagcccctg attttgcaaa cctgaagcac
2520

55 tatcaaaaac agcagagtct tccaagtta tgcagcactt ctgacccaga cacacctt
2580

ggggccccga gcactccagg gaggatctcc ctccgaatat ctgagtctgt cctgcgggac

EP 1 355 149 A2

2640

5 tccccgccac ctcatgagga ttatgaagac gaagtgtttg tgagggatcc gcacccaag
2700

gccacgtcca gcccacatt tgaacctctt cccccacccc cactcctcc accgagtcag
2760

10 gaaaccccg tglatagcat ggatgacttc cctccacctc ctcccacac tgtatgtgag
2820

gcgcagctgg acagtgagga tcccagaggg ccacgcccc gcttcaaaa actttctaaa
2880

15 gtgacaattg caagggaag gcacatgcct ggtgcagccc atgtggtagg tagtcagaca
2940

ctggcttcca gactccaaa cttctatcaag ggttcagagg ctgagtcac accacctcc
3000

20 ttcatgagcg ttcacgcccc acttgctggg tctcttggtg ggcagccagc acccatccag
3060

actcaaagcc tcagccatga tccagtcagt ggaactcagg gtttagaaaa gaaagtcagt
3120

25 cctgatcctc agaagagttc agaagacatc agaacagagg ctttggccaa ggaaattgtc
3180

caccaagaca aatctctagc agacattttg gatccagact ccaggctgaa gacaacaatg
3240

30 gacctgatgg aaggtttgtt tccccgagat gtgaacttgc tgaaggaaaa cagtgtaaa
3300

aggaaggcca tacagagaac tgtcagctct tcaggatgtg aaggcaagag gaatgaagac
3360

35 aaggaagcag tgagcatgtt ggttaactgc cctgcctact acagtgtgtc tgctccaag
3420

gctgagctac tgaacaaaat caaagagatg ccagcagaag tgaatgagga agaggaacag
3480

40 gcagatgtca atgaaaagaa ggctgagctc attggaagtc tcaccacaa gctggagacc
3540

ctccaggagg cgaaggggag cctgctcacg gacatcaagc tcaacaacgc cctgggagaa
3600

45 gagggtgagg ctctgatcag cgagctctgc aagcccaatg agtttgacaa gtataggatg
3660

ttcatagggg atttggacaa ggtggtcaac ctgctgctct cctctcggg gcgtctagcc
3720

50 cgtgttgaga atgtccttag cggccttggg gaagatgcc aataatgaaga aaggagctct
3780

ctttacgaga aaaggaagat cctggctggt cagcatgagg atgccggga gctgaaggag
3840

55 aacctggatc gcaggagcg agtagtctg ggcattctgg ccaattacct ttcagaggag
3900

cagctccagg actaccagca cttcgtgaaa atgaagtcca cgctcctcat tgagcaacgg

EP 1 355 149 A2

3960

5 aagctggatg acaagatcaa gctgggccag gagcaggtea agtgtctgct ggagagcctg
4020

ccctcagatt tcattcccaa ggctggggcc ctggctctgc ccccaaacct caccagtgag
4080

10 cccattctctg ctgggggctg tactttcagt ggtattttcc caacattaac ctctccactt
4140

taacctcttc taaaataccc aaccaaaaga tcaactgttc tctcaacact atttaactctg
4200

15 aaaaatgttt cagtacaaac cactgtttga actatctggg ttattggtgt ttgttccctga
4260

tgaaaggaaa aaaattctct ccaggaggaa gcctttttcc ttcttgccct tcctgattga
4320

20 tcttctgaga gctcgaatgc tgctggacac gtaccccttt ctattattac tttgtagtag
4380

aaagaaagtt aatgaaactg agaactgatt ggagggtgtt tgatcattta gtttttaaca
4440

25 ggctgaggca acatggatca gtgtgtgtcc ccttcaggaa tgtatccaca gtggccttcc
4500

ttgtgtgtgg gcagtgtatc ctgatggcag ggtacaagta ccattaatga agggctctgca
4560

30 acataaagcc ttaaaaagac acacactaag aaaactgtaa aaccttgaaac attgttattt
4620

atatttttta aaatggaaaa gatcactatg ttltgtgtgc taaccactta tttgattctg
4680

35 ttttgtgtg gacatagatg attacgtttg agctttgtat tttgtgaaaa ccttaatgaa
4740

atgaattcca aagat
4755

40 <210> 7
<211> 2045
<212> DNA
<213> human

45 <400> 7
gaaacttgac cccggctcat cctgtctctg gctgtggccc ggcaaagcac tgaaaacccc
60

tctggtctca gagacagtag gggcagtgcc actttctaca acctgccaac ccacacactg
120

50 gagtaattct gaaaaaatt attcctaaac tctctaagtg tggacggaga atgagcaagc
180

cccagaagta ttttacaacc agagtgggta atgaggaggg ggcttactgg aatcgtcata
240

55 tctctgaata ttgaaaacaa caactaaaaa agtggacctt ctcagaaaaa aagggcagca
300

aatgaccaag ggcgccctt ctggccgtgc ttggttgag taactgtctc tttttcccca
 360
 5 ccccatcac agggctttca gtttgcaaa ggaaaagcag ataaaaacag aacattccat
 420
 atgtttcttt ctccatcggc caaaaacatt ttgacacaat gtttgtgaaa cacctttgga
 480
 10 gaggtgcaact tctgaatgct gcctctgccg taaatcctgg ggaagggat cagcctcttc
 540
 ccaggaacca tcgccttcta taaaccgtga actcaagcag gcattttttt tttcttaccg
 600
 15 aaaggctgct atttgtcaag ggcacataat gggctctgtg ctcttattgg ctccaaatg
 660
 tgcattggcaa agagagagat gtgggcctag agcagalata ttcagcaagg tgacagcttc
 720
 20 ccataacaat tctaacactt ctatctttat gtgagaataa aatatttaag ggttgaacct
 780
 tatttlgcca aatgtatctt ttctgctttt gaattgggca gaagatttta gcaactatat
 840
 25 tctacaaatg ttacttataa cacacacaca cacatctgaa atatatgccg aaaattgacg
 900
 tctttgacct caggagagc acctgtccag gctgcctaa aggaaatggc tccagtgggt
 960
 30 ctaaacacc acatcctatc catggatagg tctagtcata acactttaga gagaatgtca
 1020
 gagcaggagg gaggaagcc gcctcttctc ggccatcaac tgcagatgat gaaagagcgg
 1080
 35 gattcaactt tgttttcttt tctgtggcc ccagtgaac ctctgccct cctgcacgt
 1140
 ctgtgtcttc atttctaaaa tgggggtgat gctttcatat tgacctcacc ccatactacc
 1200
 40 tcacagatgt gttgtgagga ttaataaaat tatgtctatg gtattttcag tttctggaga
 1260
 aaaatactta tagacagttt aactattaca tagatatata agtgaatctca gtttcttggt
 1320
 45 tgctgtgata ctaatgtgtt gttttaactt attccataaa atgacagttg tglcctagcc
 1380
 acatcagaca gctatctaag ctctggacta cccctttgtg cagctgaatc actgcagggt
 1440
 50 cgaccatgcc tgggtccaca gccatggttt ccatttctag atgaaaggat ggcctaggac
 1500
 ataggctctca aagactcttg gatcagaatc aggagattag ggaaaacagg atggatacct
 1560
 55 gagcactaac agcagtagac gtagacctct gtcctttacc atctgaggtc ttctggatc
 1620

EP 1 355 149 A2

tttgtggggt taattttgat ctgattgcat ctglttgccc ttcattcttg ttgcaagtgt
 1680
 5 gcatgggttca atccclcaca tccaggaaat gaattttgca attgggccag atgctaattt
 1740
 gcacgttgat tcaccttctt tgcctttaag cctttttttt cttttttttt ttttttgga
 1800
 10 aatgaatgta ccatttcaac tttgatttta atagtgttag ttgatattgg taataatgct
 1860
 aaccaagaga tcaatgccag atttttctct tggggtaagt tagctgaagt catttaaaga
 1920
 15 tggaaagggtg ggaaaattct ttgatatttg atgtcattgt atccacattt gttgtaagac
 1980
 atattgcata ccaattataa ttatatcaat taaagttgat aaaagcttca aaaaaaaaaa
 2040
 20 aaaaa
 2045
 <210> 8
 <211> 2096
 <212> DNA
 25 <213> human
 <400> 8
 atggagaacg agcctgtagc ccttgaggaa actcagaaga cagatcctgc tatggaacca
 60
 30 cggttcaaag tgggtgattg ggacaaggac ctggtggact ggcgaaagcc tclcctgtgg
 120
 cagggtggcc acttgggaga gaagtacgat gagtgggttc accagccggt gaccaggccc
 180
 35 atccgcctct tccactcaga cctcattgag ggcctctcta agactgtctg gtacagtgtc
 240
 cccatcatct ggggtgcccct ggtgtgtat ctgagctggc cctactaccg aacctttgcc
 300
 40 cagggcaacg tccgactctt caggtcattt acaacagagt acacgggtggc agtgcccaag
 360
 tccatgttcc ccgggtctct catgtgggg acattcctct ggagcctcat cgagtacctc
 420
 45 atccaccgct tctgtttcca catgaagccc cccagcgaca gctattacct calcatgctg
 480
 cacttcgtca tgcacggcca gcaccacaag gcacccttcg acggctcccg cctggtcttc
 540
 50 cccctgtgc cagcctccct ggtgatcggc gtcttctact tgtgcatgca gctcatcctg
 600
 cccgaggcag tagggggcac tgtgtttgcg gggggcctcc tgggctacgt cctctatgac
 660
 55 atgaccatt actacctgca ctttggctcg ccgcacaagg gctcctacct gtacagcctg
 720

aaggcccacc acgtcaagca ccactttgca catcagaagt caggatttgg taccagcact
 780
 5 aaattgtggg attactgttt ccacaccctc actccagaga aacccacct gaagacgcag
 840
 tgacaactcc caccacctcc gtectgccct cagcccgcc ctggccctt cccgaccccc
 900
 10 acccgccatt cagaccccat taagaagggt ggcttgcca ggcaggatgy gctglgtccg
 960
 gccctgcagc ctagtgaag gtgctgagg ggccctgagg caggaccgcc ctctgaccc
 1020
 15 ctggtaggag ggtcacatcc acttgggtga ggtggccctt ggtgaccac ttcttctgg
 1080
 agcgtccctg cctagagctc agccacagg actgcttcag gccgtggcca caggtagcag
 1140
 20 ccgcaagggg aatgaagaa aactgagccc tcgtggccac ctgtgtcacc cttgtgcctt
 1200
 agcctcatgg gctgcctagg agctgcctgc acggcacagc tcgttttcc acgcagaagt
 1260
 25 ggggtctgtg gatctgtgtt ccctgtcctc cctgtgttc cttctgggga ggctttggg
 1320
 gctctgaggt ggacaaagag ctctcgcaag aagagacagc gtgatgcctc ccacagtcca
 1380
 30 cccagaccc tggggcagcc cctctggccc tgccagctgc ctgcgtcgtt gggcccaggg
 1440
 tggtggcag gagtcccagc tcttgcctt aggacctggc agcttttctt gccgtccctc
 1500
 35 ccctgcctcc agaatcacag cccttctccc caaggagggc tgaggaggct tctccaccag
 1560
 tggcagcccc acccgtccc tggcattct tggcctccac cccgctcagg ccctactcg
 1620
 40 ggcgtccca gaaggagcca cctctcagt cctcacctcc ccctgcctcc cagcctccgc
 1680
 agatgagggt cctgcccctt cctcctcgt accaaaaacc tcaactgctc caggacggtc
 1740
 45 ttatttataa accagatata tgttcttagt ctggtcccag accaaggagc tggtcagacg
 1800
 gccctttcta atcctacatg ttgagcttat gtaaaaaatg ttgtttcctc ctgtttttgg
 1860
 50 ttcctttctt accacaaac cattactact tgaaactta aaaaactgcc aagtgtaaag
 1920
 gctaaagaga agcagtttga cggacctgt gatttgtact gtttctgctg gagctattta
 1980
 55 aagattttgg aataaatata caaaactacg gttgtgaaat aaaaactta attgtatatt
 2040

ttgaaaaata aaacactgaa aagaaaccaa caaaaaaaaa aaaaaaaaaa aaaaaa
2096

5

<210> 9
<211> 5640
<212> DNA
<213> human

10

<400> 9
ggaaacgcag aaaacagaga gaggcattct gagtcattctg actggatgaa gactgttcca
60

agttacaacc aaacaaatag ctccatggac tttagaaatt atatgatgag agatgagact
120

15

ctggaaccac tgcccaaaaa ctgggaaatg gcctacactg acacagggat gatctacttc
180

attgaccaca ataccaagac aaccacctgg ttggatcctc gtctttgtaa gaaagccaaa
240

20

gcccctgaag actgtgaaga tggagagctt ccttatggct gggagaaaat agaggacctt
300

cagtatggga catactatgt tgatcacctt aaccagaaaa cccagtttga aaatccagt
360

25

gaggaagcca aaaggaaaaa gcagttagga caggttgaaa ttgggtcttc aaaaccagat
420

atggaaaaat cacacttcac aagagatcca tcccagctta aagggtgcct tgttcgagca
480

30

tcactgaaaa aaagcacaat gggatttggt ttactatta ttggtggaga tagacctgat
540

gagttcctac aagtgaaaaa tgtgctgaaa gatgggtccg cagctcagga tgggaaaatt
600

35

gcaccaggcg atgttattgt agacatcaat ggcaactgtg tcctcggtca cactcatgca
660

gatgttgctc agatgtttca attggtacct gtcaatcagt atgtaaacct cactttatgt
720

40

cgtggttata cacttcctga tgacagtga gatcctgttg tggacattgt tgctgctacc
780

cctgtcatca atggacagtc attaaccaag ggagagactt gcatgaatcc tcaggatatt
840

45

aagccaggag caatggttct ggagcagaat ggaaaaatcgg gacacacttt gactggtgat
900

ggtctcaatg gaccatcaga tgcaagtga cagagagtat ccatggcatc gtcagqcagc
960

50

tccagcctg aactagtgac tatccctttg attaagggcc ctaaaagggtt tgggtttgca
1020

attgctgaca gccctactgg acagaagggtg aaaatgatac tggatagtca gtggtgtcaa
1080

55

ggccttcaga aaggagatat aatlaaggaa atataccatc aaaatgtgca gaatttaaca

1140

catctccaag tggtagaggt gctaaagcag tttccagtag gtgclgatgt accattgctt
1200

5

atcttaagag gaqgtectcc ltcaccaacc aaaactgcc aatgaaaac agataaaaag
1260

gaaaatgcag gaagtttgga ggccataaat gagcctattc ctcagcctat gccttttcca
1320

10

ccgagcatta tcaggtcagg atccccaaaa ttggatcctt ctgaggtcta cctgaaatct
1380

aagactttat atgaagataa accaccaaac accaaagatt tggatgtttt tcttcgaaaa
1440

15

caagagtcag ggtttggcct caggggtgcta ggaggagatg gacctgacca gtctatatat
1500

attggggcta ttattccctt gggagcagct gagaagatg gtcggctccg cgcagctgat
1560

20

gaactaatgt gcattgatgg aattcctgtt aaagggaaat cacacaaaca agtcttggac
1620

ctcatgacaa ctgctgctcg aaatggccat gtgttactaa ctgtcagacg gaagatcttc
1680

25

tatggagaaa aacaacccga ggacgacagc tctcaggcct tcatttcaac acagaatgga
1740

tctccccgcc tgaaccgggc agaggtecca gccaggcctg caccocagga gccctatgat
1800

30

gttgtcttgc aacgaaaaga aaatgaagga ttgggctttg tcatcctcac ctccaaaaac
1860

aaaccacctc caggagttaa tcttcataaa attggccgag tcatagaagg aagtccggct
1920

35

gaccgctgtg gaaaactgaa agttggagat catatctctg cagtgaatgg gcagtccatt
1980

gttgaactgt ctcatgataa cattgttcag ctgatcaaag atgctgggtg caccgtcaca
2040

40

ctaacggtea ttgctgaaga agagcatcat ggtccaccat caggaacaaa ctcagccagg
2100

caaagcccag ccctgcagca caggcccatg ggacagtcac aggccaacca catacctggg
2160

45

gacagaagtg ccctagaagg tgaaattgga aaagatgtct ccacttctta cagacattct
2220

tggtcagacc acaagcactt tgcacagcct gacaccgcag taatttcagt tgtaggcagt
2280

50

cggcacaatc agaaccttgg ttgttatcca gtagagctgg agagaggccc ccgyggcttt
2340

ggattcagcc tccgaggggg gaaggagtac aacatggggc tgltcacct tegtcttgc
2400

55

gaagatggtc ctgccatcaa agatggcaga attcatgttg gtgaccagat tgttgaaatc

EP 1 355 149 A2

2460
aatggggaac ctacacaagg aatcacacat actcgagcaa ttgagctcat tcaggctggt
2520
5
ggaaataaag ttctttcttct tttagaggca ggaactggct tgatacctga ccatgggttg
2580
gctccttcg gtctgtgtc ctacgtgaaa cccgagcaac attaaggctt tcagggtttt
2640
10
tcttggtctt tccttaaaaa gacttggtga ttgggatatt aataatcctt cgtcttcaaa
2700
tgtgatttat gatgaacagt caccattacc cccatcttca cattttgctt ccatatttga
2760
15
agagtctcac gtgccagtaa ttgaagaatc ttlgagagtt cagatatgtg aaaaggcaga
2820
agaattaaag gacattgtgc ctgaaaagaa aagcacttta aatgaaaatc agcctgagat
2880
20
aaagcatcag tctcttctcc agaaaaatgt gagtaagagg gatccacca gcagtcatgg
2940
gcacagtaac aagaaaaatc tattaanaat agaaaatggt gttacacgaa gagglagatc
3000
25
ggttagtccc aaaaagccag ccagtcaaca ttcagaggaa catttgata agattcctag
3060
tccttcaaaa aataacccca aaagaagacc cagagatcaa tccctcagcc ccagcaaagg
3120
30
ggaaaataaa agttgtcagg tcagcaccag ggcaggctct ggacaagatc agtgcagaaa
3180
aagcagaggt cggtcggcca gcccaaaaaa gcagcaaaaa allgaaggaa gcaaagctcc
3240
35
atcaaatgct gaggccaaat tattagaggg taagagtcga agaatagcag gctatacggg
3300
cagtaatgct gaggagatcc cagatgggaa ggaaaaatca gacgtcatca ggaaagatgc
3360
40
aaagcagaat cagttggaaa aaagcagaac aaggtctcca gagaaaaaaa tcaaaagaat
3420
ggttgagaaa tctcttccat ccaaatgac taataagact acaagtaaag aagtatctga
3480
45
aatgaaaaa ggaaagaaag taaccacagg agaaacaagt tctagtaacg ataaaatagg
3540
agaaaatgct cagctatcag aaaagaggct gaagcaagaa cctgaagaga aggtagtctc
3600
50
aaacaaaaca gaagatcaca aagggaaga actagaggca gctgacaaaa acaaagagac
3660
tggaaggctt aaaccggaaa gcagttctcc agttaagaaa aactgataa ctccagggcc
3720
55
ctggaagggt ccaagtggaa ataaagtcac aggcactall ggtatggctg agaaacggca

EP 1 355 149 A2

3780

5 gtaaccttta gtataaaaca aagaaaaaca agttgtaatc ttttcttaca gcagcathtt
3840

tccagaaaaa gccttttttt ttttttcaga tatlctgaaa cagataagta calgttaatg
3900

10 tgagcctcaa gtlacctagg ctgcatgaag ggcctttagg attgctaaga accaactgtc
3960

ccccggccg gctgccctcc ctgctctca ggaaggagct gcattccat gctcatctga
4020

15 ccgcccctgc tcaggctgcc cagctcgtct tcatgagtgt ctgaacaaat gacatatgtt
4080

gatattaaca atgtgggtcac aactcacttt gtatttgtgc caagttatct actgtatcat
4140

20 gtctgttttt atcctttttg ttcagctgtt tccacagtaa tgaaaaagtt aggtttggct
4200

tggaagttga tgatctcaat agcatgttgc atgtttacag agagaaatat gtgagtcctt
4260

25 gcagaagaag agactgttaa ctcatcgtta aagatggccg ttgtctcttc taacagctac
4320

tgatgatgtc ccactttaaa aataaaaccc ccaaacatca ctactttaag gaaaaaaaaa
4380

30 atgtagtcca atattgatgc tttcttatgg ctttttattt taatttggct ggataagttg
4440

tttcaaataa ctgttaaaga tattacttac aattgaatgt ttgaaataag aaagtacttt
4500

35 aagcaataga gttcatctcc tgctgtgtta tccaacctcg atglalactt acagcatctc
4560

aggtcaccc ttttatttca gttatttaat tatgaaacca taaagaagca tgtggaaata
4620

40 gtgtttattg ctctttgaag aaaaaccacc aactatttct ggatattttg gctgtaccta
4680

ctactaaagt cattagtctt taatacataa tacatatattg aaaagtaaac atattatata
4740

45 gattatgtga gggacttaat catgaaacca gtttcacagt ccaagtacca actcttctgg
4800

tagcaggtgc acaagcttgg gtgttttaaaa acaacctgtg tagggatatgc ccagcaaag
4860

50 aggacaaatg tgtagacagt acttactgga tcttatttaa cttttagcta cattaactaa
4920

ctttcttatt taaaaacaag aaaggagac taaacatctg cttaacttgt acacattttc
4980

55 agaattcttt ttaaaagtct agttaaatgt gtttcttaga agttggagac tgtaacaac
5040

ttccataaaa tagatccagg tttttcagtt cctgaagca gcattcagta gcattctat

5100

aaataaaggc accttctgag aataaaacta ttttatggag tgtgtgaaca cacttgttct
5160

5

gtcacctggg ttcattctgt tgtgaagcac attaggtcca ggctcttccc tctgggaglc
5220

tgactgtgaa actcttllaac ccaacaactc aattagcccc thtagataag acatgcttcc
5280

10

cagagtgaga tttttgaaat ccccttttca tccagaacta tatttaccba cctattgtaa
5340

ctattcaaat agagcaaaat taggaggctt gataaatact aagaatttag taccacagaa
5400

15

attatatttatt attttccctg tagtcacaaa ttagtgataa cgaatcctat ttttgtaa
5460

tgtgacataa ctttgatgtc atatgttgtc ctatgtgggt cttcctaagt aaactctgta
5520

20

ctgattatat actgacttag caatgtggcc ttggaatgct gagcaaaatg tggatgtact
5580

ggttgtaa at gtttatatat tgtacagtac ctttatatat acacttgagg ttctgattag
5640

25

<210> 10
<211> 457
<212> DNA
<213> human

30

<220>
<221> misc_feature
<222> (242)..(242)
<223> any kind of base

35

<220>
<221> misc_feature
<222> (369)..(369)
<223> any kind of base

40

<220>
<221> misc_feature
<222> (394)..(394)
<223> any kind of base

45

<220>
<221> misc_feature
<222> (406)..(406)
<223> any kind of base

50

<220>
<221> misc_feature
<222> (457)..(457)
<223> any kind of base

55

<400> 10

EP 1 355 149 A2

tcagtcactc tttcacctcg ccaagcttc actgacctac tgattgaatt gtatgtgaga
 60
 5 aataaaatgt catcatatta agccactggg atttgtatgt ttatctgtta tagcagcaag
 120
 tcttaattta cctaatacac acattgtgac agatgttctt aatgtccac cccatattgt
 180
 10 tacatgtcca gctttgagga tccctggcat gtgggggtag gagtttctgg gcatgctgga
 240
 tncaattccc acttttaagg catctgtggc ctctgtggcc tctgtggcct teactgttat
 300
 15 ggaagggatt tatctggggc accataggaa actttacat ggcacagtgg acaacctagg
 360
 aggggggtgng gaggaggggc cttcaggccc aacngggggg accagngttc gtgggggttag
 420
 20 ggtgggttgg ggggttttcc ctcttaccgg tgggggn
 457
 <210> 11
 <211> 1493
 <212> DNA
 25 <213> human
 <400> 11
 aatagggttg gcggtgcag cgggaggcaa acagcccgcc cggcaccacc atgctcgccc
 60
 30 tggaggctgc acagctcgac gggccacact tcagctgtcl gtaccagat ggcgtcttct
 120
 atgacctgga cagctgcaag cattccagct accctgattc agagggggct cctgactccc
 180
 35 tgtgggactg gactgtggcc ccacctgtcc cagccacccc ctatgaagcc ttcgaccggg
 240
 cagcagccgc ttttagccac cccaggetg cccagctctg ctacgaacc cccacctaca
 300
 40 gccctgcagg gaacctcgaa ctggccccca gcctggaggc cccggggcct ggccctccccg
 360
 cataccccac ggagaacttc gctagccaga ccctggttcc cccggcataat gccccgtacc
 420
 45 ccagccctgt gctatcagag gaggaagact taccgttggg cagccctgcc ctggaggtct
 480
 cggacagcga gtcggatgag gccctcgtgg ctggccccga ggggaaggga tccgaggcag
 540
 50 ggactcgcaa gaagctgcgc ctgtaccagt tcctgctggg gctactgacg cgcggggaca
 600
 tgcgtgagtg cgtgtggtgg gtggagccag gcgccggcgt cttccagttc tcctccaagc
 660
 55 acaaggaact cctggcgcgc cgctggggcc agcagaaggg gaaccgcaag cgcattgacct
 720

EP 1 355 149 A2

accagaagct ggcgcgcgcc ctccgaact acgccaagac cggcgagatc cgcaaggtea
780

5 agcgcaagct cacctaccag ttcgacagcg cgctgctgcc tgcagtcncc cgggectgag
840

cacacccgag gctcccacct gcggagccgc tgggggacct cacgtcccag ccaggatccc
900

10 cctggaagaa aaagggcgtc cccacactct aggtgatagg acttacgcat cccaccttt
960

tggggtaagg ggagtgtgc cctgccataa tccccaagcc cagcccgggc ctgtctggga
1020

15 ttcccactt gtgcctgggg tccctctggg atttcttgt catgtacaga ctccctggga
1080

tcclcatgtt ttgggtgaca ggacctatgg accactatac tcggggaggc agggtagcag
1140

20 tgcttcaga gtccaagag ctctctggg atttcttgt gatctctgat tccccagtga
1200

ggcctgggac ctttttaaga tcgctgtgtg tctgtaaacc ctgaatctca tctggggtgg
1260

25 gggccctgct ggcaaccctg agccctgtcc aagggtccct cttgtcagat ctgagatttc
1320

ctagttatgt ctggggccct ctgggagctg ttatcatctc agatctcttc gccatctat
1380

30 ggctgtgttg tcacatctgt cccctcattt ttgagatccc ccaattctct ggaactattc
1440

tgctgcccct ttttatgtgt ctggagttcc ccaatcavat ctagggtcc tcc
1493

35

<210> 12
<211> 2292
<212> DNA
<213> human

40

<400> 12
ccatgggttc cccctcagcc tgtccatata gagtgtgcat tccctggcag gggctcctgc
60

tcacagctc gcttttaacc ttctggaacc tgccaaacag tgcccagacc aatattgatg
120

45

tcgtgccgtt caatgtcgca gaagggaagg aggtccttct agtagtccat aatgagtcac
180

agaatcttta tggctacaac tggtaaaaag gggaaagggt gcatgccaac tatcgaatta
240

50

taggatatgt aaaaaatata agtcaagaaa atgccccagg gcccgcacac aacggtcgag
300

agacaatata ccccaatgga accctgctga tccagaacgt taccacaat gacgcaggat
360

55

tctataccct acacgttata aaagaaaatc ttgtgaatga agaagtaacc agacaattct

420

5 acgtattctc ggagccaccc aagccctcca tcaccagcaa caacttcaat ccggtggaga
 480
 acaaagatat tgtggtttta acctgtcaac ctgagactca gaacacaacc tacctgtggt
 540
 10 gggtaaacaa tcagagcctc ctggtcagtc ccaggctgct gctctccact gacaacagga
 600
 ccctcgttct actcagcgcc acaagaatg acataggacc ctatgaatgt gaaatacaga
 660
 15 acccagtggg tgccagcgc agtgaccag tcaccctgaa tgtccgctat gagtcagta
 720
 aagcaagttc acctgacctc tcagctggga ccgctgtcag catcatgatt ggagtactgg
 780
 20 ctgggatggc tctgatatag cagccttggg gtagtttctg ctttcggga agagtgttt
 840
 tattatccac ctgcagactg gactggattc ttctagctcc ttcaatccca ttttctcctg
 900
 25 tggcatcact aagtataaga cctgctctct tctgaagac ctataagctg gaggtggaca
 960
 actcaatgta aatttcaagg aaaaaccctc atgcctgaga tgtggggccac tcagagctaa
 1020
 30 ccaaagtgt caacaccata actagagaca ctcaaattgc caaccaggac aagaagtga
 1080
 tgacttcatt ctgtggacag tttttccaa gatgtccaa gcctcatcgt gacgaggctc
 1140
 35 ttatccact ccattttcc ctgctcatgc ctgcctcttt aatttggtta gataatgctg
 1200
 taactagaat ttcacaatca gcgccttggt caggcaattt gacagagtgt tggatgtgtc
 1260
 atgtcatcat gtcaaaccac aatatttgac ctaagggatc ctttattctg cccagtggct
 1320
 40 aactttaaca acatccctaa tacaactgtt tattcaaag cacggtggc cctgttagag
 1380
 ttagacctct agactcacct gttctcacgc cctgttttaa ttaaccag ctatgggatg
 1440
 45 ccagataaca gaattgctgc ctacgagctg aacagggagg agtttgtgca gttgctgaca
 1500
 cttcttggtg cacataaata aatacagtgg gtactataga gactcagttg caaaaattaa
 1560
 50 caaatatgct gcttgattaa aatgggtagg cttctcatgt ggctcattct ttaatctatt
 1620
 ctcttttatt tggtttggt catggggtct ctgcctatgg atcatacttc aaactcttg
 1680
 55 tgtgactctc ctgattgtca caatattagt taccctgggt tgctgtattc tctaaaacct

1740

5 ttaaattgttt gcatgcagcc attcgtcaaa tgtcaaatat tctctctttg gctggaatga
1800

caaaaaclca aataaatgta tgattaggag gacatcataa cctatgaatg atggaagtcc
1860

10 aaaatgatgg taactgacag tagtgtaaat gccttatgtt tagtcaaact ctcatttagg
1920

tgacagcctg gtgactccag aatggagcca gtcattgtaa atgccatata ctcacactga
1980

15 aacatgagga agcaggtaga tcccagaaca gacaaaactt tcctaaaaac atgagagtcc
2040

aggtgtctg agtcagcaca gtaagaaagt cctttctgct ttaactctta gaaaaaagta
2100

20 atatgaagta ttctgaaatt aaccaatcag tttatttaaa tcaatttatt tatattcttc
2160

tgttctgga ttcccathtt acaaaacca ctgttctact gttgtattgc ccagtaggag
2220

ctatcactat attttgcaga atggaaactg ccctgactct tgaatcaca aaaaaagcca
2280

25 attgtatctg tt
2292

30 <210> 13
<211> 519
<212> DNA
<213> human

<220>
35 <221> misc_feature
<222> (212)..(212)
<223> any kind of base

<220>
40 <221> misc_feature
<222> (451)..(451)
<223> any kind of base

<400> 13
45 gaaacaacaa cagtgtaatc ttaacaggg atgttaaagg taagaagtca ggaagataaa
60

ccaaaatgat tgagtatgat aaagaatttt gcatggcgat taaaatagaa aacctataaa
120

50 tgtagaaaaa gcagggtctgg acttagcaaa gaaacaatat agtttggaga aggcattgaaa
180

taagttcttt tcatgttcac tgctggtcac ancataacag agagtgatgt ggagagcttt
240

55 gggaagggtt cacgttgagt tacatcagtg gtcaacaatg gagcaacaag actccgtaga
300

EP 1 355 149 A2

ggatgccacc ctgggagaaat tgcaggagaa aggaggctga agcacaactg gtaatagcct
 360
 5 tcatatattt aatggatatg caaataaagc tctgattaat tgtattttca cttattatat
 420
 atcatctttg gacctttcta aaagtgggac nctagaaaag atatactgaa actccaaaag
 480
 10 aatacttcag ctcgagttga atggattcaa gatgttggt
 519
 <210> 14
 <211> 5294
 <212> DNA
 15 <213> human
 <400> 14
 ggctcgcatc cccatagtcg tgggttacag tgaaggtagc ccccgcgctc tgctctggag
 60
 20 aggcagggtg ggataggga cgtctcgagt ggcgcccga gtcaggttg tgctctggag
 120
 ccgcccgcctc cggcgctcc tagggctgtt taagaagaag ggctctgcca aggctgagaa
 180
 25 tgacaaacat ctaagttagt ggcctggcca ggggccaggg tctgcagtgg atgagacca
 240
 ggacaacgtc ttctttccca gtgggagacc ccccccctg gaagagctgc acactcaggc
 300
 30 ccaggagggtg ctccgctccc tacaacacca agagaaacag aaactgaaca aggggtggctg
 360
 ggaccatgga gacaccaga gtatccagtc ctcccgagc gggccggatg aagacaacat
 420
 35 ctcttctgct agtcagacca catctacgt ggctgagagc tccacagcag aggcgcgct
 480
 ctccatccgc tcggagatga tccagcgcaa aggtccacc ttccgacccc atgactcatt
 540
 40 tcccaaactc ggaaagtcag ggcgycgtcy gcgggagcgg cggagcactg tgctgggact
 600
 cccgcagcat gtgcagaagg agcttggcct gaggaatgag cgtgaggcac caggcacgcc
 660
 45 ccgggctcct ggtgcacggg atgccgtacg catccccaca gtggacggcc gccccgagg
 720
 cacctcaggg atgggggccc ggggtgtccct gcaggcgctg gaggcgagg cagaggctgg
 780
 50 cgctgagaca gaggccatgc tgcagcgcca cattgaccgt gtctaccggg atgacacctt
 840
 tgttggccgg tccacgggta cccgggcccc accattgacc cggcccatgt cctagcagt
 900
 55 gcctggattg acaggagggg cagggcctgc agagcccctg agcccggcca tgtccatctc
 960

cccccaggcc acctacctgt cgaagttgat tccacatgct gtgctgccgc ctacagtggg
 1020
 5 cgtgggtggc ctaggccgct gcagcctgcg cacactaagc cgctgcagcc tgcactcggc
 1080
 cagcccagcc tcagtcgcgt cgctggggcg cttctcctcc gtctccagcc cacagccccg
 1140
 10 cagccgccac ccctcctcct ccagtgcacac ctggagccac tctcaatcct ccgacaccat
 1200
 tgtgtctgac ggttcacccc tctcctctaa ggggtggctct gagggccagc cggagagctc
 1260
 15 tacggctagc aatagcgtgg taccctcctc ccagggaggc agtgggaggg gctctcccag
 1320
 tgggggcagc actgctgagg cctcagacac actcagcatt cygagcagtg ggcagttgtc
 1380
 20 tggccggagt gtgtccctgc gtaagctgaa gcggcctcca cccctcccc gccggacca
 1440
 ctccctccat cagcggggct tagcagtgcc tgatgggcca ttaggggttc cccctaagcc
 1500
 25 tgagcgtgaa cagcagcccc agctgcctcg gccaccacc actggtggct cagaaggggc
 1560
 gggggcagca ccctgtccac ccaaccagc caacagctgg gtacctggct tgtctccggg
 1620
 30 tggttcccg cgcctccac ggtcccca gaacgacactt tcgcccctca gtggatactc
 1680
 gagccaaagt ggtactcca ccctcctcct caagggcctg gcaggtcccc ctgcttcccc
 1740
 35 aggcaaggcc cagcccccta aaccagagcg tgtcacgtct ctctgctccc ctggggcctc
 1800
 cgtctcctct tccctcacgt ctttatgttc ctctcctct gaccagccc cctcagaccg
 1860
 40 ctctgggcca cagatattga cccctcggg tgacaggttt gtcatactc ctcaccccaa
 1920
 ggtgcctgcc ccttctccc cactcctc caagcccagg agccctaacc cagctgcccc
 1980
 45 tgcctlagcc gccctgctg tggttcctgg gcctgtttct accactgacg ccagtcctca
 2040
 gtcccctccc actccccaga caacctgac tccactgcag gagtctctg tcatctccaa
 2100
 50 agaccagtca ccccaacctt cccaccccc atcttatcat ccacccccc caccactaa
 2160
 gaagccagag gtggttgtgg aggcaccatc tgcctcagag actgctgagg agcccccca
 2220
 55 agatcccaac tggccccctc cccaccccc tgcctcagag gagcaggacc tgtccatggc
 2280

tgacttcccc ccaccagagg aggetttttt ctctgtggcc agccctgagc ctgcagggcc
 2340
 5 ttcaggetcc ccagagcttg tcagctcccc ggctgcttcg tctctctcag ctactgcttt
 2400
 gcagattcag cccccgggta gcccagaccc tcttcagct ccgccagccc cagctcctgc
 2460
 10 tagttccgcc ccaggcatg tggccaagct cctcagaag gaaccgggtg gctgtagcaa
 2520
 ggggtggtgg cctcccaggg aggacgtagg tgcgccctg gtcacgccct cgtcctgca
 2580
 15 gatggtgcgg ctgcgtccg tgggtgctcc aggaggggt cccaccccag cactggggcc
 2640
 atcggccccc cagaaaccac tgcgaagggc cctgtcaggg cgggccagcc cagtgcctgc
 2700
 20 cccctctca gggctccatg ctgcggtccg actcaaggcc tgcagcctg ccgccagtga
 2760
 aggctctca agtgcctcag ccaacggacc gcctgaggca gagccacggc ctccccagtc
 2820
 25 cctgcctca acggccagtt tcatcttctc caagggtct aggaagctgc agctggagcg
 2880
 gcccggtgcc cctgagaccc aggetgacct ccagcgaat ctggtggcag aactccggag
 2940
 30 catctcagag cagcggccac ccaggcccc aaagaagtca cctaaggctc cccacctgt
 3000
 ggcccgaag ccgtctgtg gagtcccccc accgcctcc cccagttacc ctgagctga
 3060
 35 gcccttact gtcctccca ccaatgggt cctcacacc caggacagga ctaagagga
 3120
 gctggcggag aatggaggtg tctgcagct ggtgggcccc gaggagaaga tgggcctccc
 3180
 40 gggctcagac tcacagaaag agctggcctg accaccaggc acctcactgg cactgctgac
 3240
 ccattccaga aacacaatct cagggacctg agcagctcca aggacgagag gatacagcag
 3300
 45 acacaaccta atagagaggg cgctgcagc cttaacctcc acggccttcg atacttatgc
 3360
 aagcctggtg ttgctcctgt cctcagagtc atcctgcgct catgcctttt cccgaatgg
 3420
 50 ttcacctctg gcagttgccg cttcagtctt ggccttagcc tcatcttgaa gtgggtagct
 3480
 ggcgggagag ggtggctgcg cccctgctg gccctgaggc tgcagagttg ggagcaggac
 3540
 55 acctcacctg agtttcattt ttttcatgt ccaaaccatg cacatactat agtcagaat
 3600

caaagcactt ttgaaaagtg gctgcatggc catcctccag ggcccaggaa gttgcattcc
 3660
 5 aagggcctgt ttacatggca gcagaatcca tccccggcag tcagcccata gcttgggacc
 3720
 agtctgtgcc ctctgcccc gtccagttta ctctcttgg ttcctgaagg tggccaagtc
 3780
 10 attgtgttcc cacaggcttc tctaggctgg gggcagggtg ggggctgtgg aattccaaag
 3840
 cacaaaaggt gcagagggga ttggccttcc tgtgcctcaa ctcaccaacc accctcctgc
 3900
 15 cttccagttc tgccagggtc tccatgctgg ggacaagtag gagactgccca gggcccaaag
 3960
 aaalgggtga gcagtagagt catctcgggg cacttggcag tgtcaagcac ctgccccttg
 4020
 20 cctccttgac cacactgggg tgggtggggc ccagcactt cagaggcagg agcctttggg
 4080
 ctgagcaagc actgaggagg tggatggaag ggagcatctg gaggggggga gcttccttga
 4140
 25 gcagtgggcc caggcctggc cctccacact tcattctctg acctttctct ctctcatitt
 4200
 cggtgcatgt cttttctgca gctgccttcc agcacagggt gttccactgg gggcagctaa
 4260
 30 cgctgagtga caaggatggg aagccacagg tgcattttac tcaagtcttc tctagtcaat
 4320
 gaggggcacc cagtgttctt agggcaggct ggggtgggtgt cccctaggta tcagcctctc
 4380
 35 ttactgtact ctccgggaat gttaaccttt ctattttcag cctglgccac ctgtctaggc
 4440
 aagctggctt ccccatggc ccctgtgggt ccacagcagc gtggctgccc ccaggggcca
 4500
 40 ccgcttcttt ctgtatctc ttctctaac agtgacttgg gcttgagtct ggcaaggaa
 4560
 cttgctttta gttcaccac caaggagaga ggttgacatg acctccccgc cccctcacca
 4620
 45 aggctgggaa cagaggggat gtggtgagag ccaggttcct ctggccctct ccagggtgtt
 4680
 ttccactagt cactactgtc ttctccttgt agctaaalcaa tcaatattct tcccttgctt
 4740
 50 gtgggcagtg gagagtgtg ctgggtgtac gctgcacctg cccactgagt tggggaaaga
 4800
 ggataatcag tgagcactgt tctgtctcaga gctcctgac taccacccc cctaggatcc
 4860
 55 aggactgggt caaagctgca tgaaaccagg ccctggcagc aacctgggaa tggctggagg
 4920

tgggagagaa cctgacttct ctttcctctt ccttcctcca acattactgg aactctatcc
 4980
 5 tggtaggac ttctgagctt gtttcctgc tgggtgggac agaggacaaa ggagaaggga
 5040
 gggcttagaa gaggcagccc ttctttgtcc tctggggtaa atgagcttga cctagagtaa
 5100
 10 atggagagac caaaagcctc tgatttttaa ttccataaa atgttagaag tatatatata
 5160
 catatatata ttcttttaaa tttttgagtc tttgatatgt ctaaaaatcc attccctctg
 5220
 15 ccctgaagcc tgagtggagac acatgaagaa aactgtgttt catttaaaga tgtaattaa
 5280
 atgattgaaa cttg
 5294
 20
 <210> 15
 <211> 988
 <212> DNA
 <213> human
 25
 <400> 15
 gtcgtgaggc gggccttcgg gctggctcgc cgtcggtgc cggggggttg gcctgggtgt
 60
 cattggctct gggaagcggc agcagaggca gggaccactc ggggtctggt gtcggcacag
 120
 30 ccattggcggg cgcgttggtg cggaaagcgg cggactatgt ccgaagcaag gatttcggg
 180
 actacctcat gactacgcac ttctggggcc cagtagccaa ctggggctctt cccattgctg
 240
 35 ccattcaatga tatgaaaaag tctccagaga ttatcagtgg gcggatgaca ttgcccctt
 300
 gttgctattc ttgacattc atgagatttg cctacaaggt acagcctcgg aactggcttc
 360
 40 tgtttgcatg ccaagcaaca aatgaagtag ccagctcat ccaggagagg cggcttatca
 420
 aacacgagat gactaaaacg gcatctgcat aacaatggga aaaggaagaa caaggtcttg
 480
 45 aagggacagc attgccagct gctgctgagt cacagatttc attataaata gcctccctaa
 540
 ggaaaataca ctgaatgcta tttttactaa ccattctatt tttatagaaa tagctgagag
 600
 50 ttctaaacc aactctctgc tgccttaca gtattaaata tttacttct ttccataaag
 660
 agtagctcaa aatatgcaat taatttaata atttctgatg atgttttata tgcagtaata
 720
 55 tgtatatcat ctattagaat ttacttaatg aaaaactgaa gagaacaaaa tttgtaacca

EP 1 355 149 A2

780

ctagcactta agtactcctg attcttaaca ttgtctttaa tgaccacaag acaaccaaca
840

5

gctggccacg tacttaaaat ttgtcccca ctgtttaaaa atgttacctg tgtatttcca
900

tgcagtgtat atattgagat gctgtaactt aatggcaata aatgatttaa atatttgtaa
960

10

aaaaaaaaa aaaaaaaaaa aaaaaaaa
988

<210> 16
<211> 4908
<212> DNA
<213> human

15

<400> 16
ggataacctc gcagggtggg ccggaggggc ggccgccc ctgcctgtgc lgcggcgatg
60

20

gcccagtgtg tacaatcagt gcaggagcta atcccgact ccttcgtccc ctgtgtcgtc
120

gcgctgtgca gcgacgaagc cgagcggtc actcgtctca atcacctcag cttcgcgag
180

25

ctgcttaagc ccttctccc cctcacttcc gaggttcaca tgagagatcc taataatcaa
240

cttcacgtaa ttaaaaattt gaagatagca gtaagcaaca ttgtcaccca gccacctcag
300

30

cctggagcca tccggaagct tttgaatgat gttgtttctg gcagtcagcc tgcagaagg
360

ttagtagcta atgtgattac agcaggagat tatgacctta acatcagtgc cactactcca
420

35

tggtttgagt cttacagaga aacctttctt cagtcgatgc cagcatogga tcatgaattt
480

ctgaaccact atttagcatg tatgttggtg gcgtcatcta gtgaagctga acctgtggaa
540

40

cagttttcaa agttgtcaca agaacagcat cgaattcagc acaacagtga ttattcctac
600

cccaagtggg ttataccaaa tacacttaaa tactatgtac tttacatga tgtaagtga
660

45

ggagatgaac agagagctga atcaatttat gaagaaatga aacagaaata tggaactcag
720

ggttgctatt tacttaaaat taattctcga acatctaata gagcatcaga tgaacagata
780

50

ccagatcctt ggagtcagta tctccagaaa aatagtattc aaaaccagga atcatatgaa
840

gatggccctt gtactataac ttcaaataag aattctgata ataacttgct ttcattggat
900

55

EP 1 355 149 A2

ggattagata acgaagtcaa agatggctta ccaaataact ttagagctca cccacttcag
 960
 5 ttggagcaat ccagtgaccc ttctaacagt attgatggcc cagatcatct aagatctgct
 1020
 tcacggttac atgaaacaaa gaaaggaaat actggaataa ttcattggtgc atgtttaaca
 1080
 10 cttactgac atgatagaat tcgacagttt atacaagagt ttcacatttcg gggccttttg
 1140
 ccacatatag agaaaacaat taggcaatta aacgatcagc taatatcaag aaaagggttg
 1200
 15 agtcgatctc tattttctgc aactaaaaaa tggtttagtg gcagtaaagt tccagaaaag
 1260
 agcattaatg acctgaaaaa tacatctggc ttgctgtatc ctccggaagc accagaactt
 1320
 20 caaatcagga aaatggctga cttatgtttt ttggtgcagc attatgattt ggcttacagt
 1380
 tgctatcata ctgcaaagaa agattttctt aatgatcaag caatgcttta tgcagctggt
 1440
 25 gccttggaag tggcagcagt gtctgctttt cttcaaccag gagcacctag gccatatcct
 1500
 gctcattaca tggatacagc aattcagaca tacagagata tctgcaagaa tatggtgttg
 1560
 30 gctgaaagat gtgtgttgc tagtgctgaa cttttaaaaa gccaaagcaa atattcagag
 1620
 gctgcagctc tctaataacg gttgaccagt gaggattctg atcttcgaag tgcacttctt
 1680
 35 ttggaacagg cagcacattg ctttataaac atgaaaagtc ccatgggttag aaaatatgca
 1740
 tttcatatga tattggcagg ccatcgattt agtaaagcag ggcagaaaaa gcatgcttta
 1800
 40 cgctgttatt gtcaagccat gcaagtttac aaaggaaaag gctgggtctt tgcagaggat
 1860
 cacattaatt tcaactattg gcgccagtcc tatactctta gacagctgga taatgctgtg
 1920
 45 tctgctttta ggcatattct aattaatgaa agtaaacaat ctgctgctca acagggggct
 1980
 ttctcagag aatatcttta tgtttacaag aatgtaagtc agctgtcacc agatggctct
 2040
 50 ttgccacagc ttcttttacc gtatattaac agttcagcaa caggggtttt ttttgccat
 2100
 gacagacgac cagcggatgg tgaaaaacaa gcagctactc atgtaagtct tgatcaagaa
 2160
 55 tatgallctg aatcctctca gcagtggcga gaacttgagg aacaagttgt ttctgtggtt
 2220

EP 1 355 149 A2

aacaaaggag taattccatc caattttcat cccacacaat actggttgaa cagttactca
 2280
 5 gataatlcaa gatttccact tgcagttgta gaagaaccaa ttacagtga agtggctttt
 2340
 agaaaccctt tgaaagttct acttttgtg actgatttgt cattgctttg gaagtttcat
 2400
 10 cctaaagatt tcagtggaaa ggataatgaa gaagttaaac aactagttac aagtgaacct
 2460
 gaaatgattg gagctgaagt tatttcagag ttcttaalta atggcgaaga atcaaaagtg
 2520
 15 gcaagactaa agctctttcc ccatcacata ggggagctgc atattctggg agttgtttat
 2580
 aatcttgga ctattcaggg ctctatgaca gtagatggca ttggtgctct tcccgatgt
 2640
 20 cacacaggaa aatattcctt gagtatgtca gtccgaggga agcaggattt agaaattcaa
 2700
 ggtcctcgac tlaacaacac aaaagaagag aaaacatctg tlaaatatgg cctlgatcga
 2760
 25 cgttttagatc ccataatcac agaagaaatg ccaactgttg aggtgttctt tatacatttt
 2820
 cctacagggc ttctctgttg agaaatccga aaagcatatg tagaatttgt caatgtcagc
 2880
 30 aaatgtccac ttactggatt gaaggttgtt tctaaacgtc cagagttctt tactttcgg
 2940
 ggtaatactg ctgttctaac accactaagt cctcagctt ctgagaattg tagtgcttac
 3000
 35 aagactgttg tgacagatgc tacctctgtg tgtacagcac tcatatcatc agcttcttct
 3060
 gtagactttg gcattggcac aggaagtcaa ccagaggtga ttctgttcc ccttcctgac
 3120
 40 actgttcttc taccgggagc ctcaagtgc ctgccaatgt ggttacgtgg gcctgatgaa
 3180
 gaaggtgtcc atgaaattaa ctttttgttt tactatgaaa gtgtcaaaaa gcagccaaaa
 3240
 45 atacggcaca gaatattaag acacactgca attatttgta ccagtcgggc tttaaatgta
 3300
 cgggccactg tctgcagaag taattctctt gaaaatgaag aaggcagagg aggcaatatg
 3360
 50 ctagtctttg tggatgtgga aaataccaat actagtgaag caggcgtaa ggaattccac
 3420
 atagtcaag tatcaagtag tagcaaacac tggaggttac agaaatctgt aaatctttct
 3480
 55 gaaaacaaag atgccaaact tgccagtagg gagaaggga agttttgctt taaggcaata
 3540

EP 1 355 149 A2

agatgtgaga aagaagaagc ggccacacag tcctctgaaa aatatacctt tgcagatata
 3600
 5 atctttqgaa atgaacagat aataagttca gcaagcccat gtgcagactt cttttatcga
 3660
 agtttatctt ctgaattgaa aaaaccacaa gctcacttgc ctgtgcatac agaaaaacag
 3720
 10 tcaacagagg atgctgtgag attgattcaa aaatgcagtg aggtagattt gaatattgtc
 3780
 atattatgga aggcatacgt tgtggaagac agtaaacagc ttattttgga aggtcaacat
 3840
 15 catgttaltc ttcgcactat aggaaaagaa gccttttcat atcctcagaa acaggagcca
 3900
 ccagaaatgg aactattgaa atttttcagg ccagaaaaca tlacagtttc ctcaaggcca
 3960
 20 tcagtagagc agctttctag tctcattaaa acgagtcttc actaccaga atcatttaat
 4020
 catccatttc atcaaaaaag cctttgttta gtaccagtca ctcttttact ttccaattgt
 4080
 25 tctaaggctg atgtagatgt catagttgat cttcggcata aaacaacaag tccagaagca
 4140
 ctggaaatcc atggatcatt cacatggctt ggacaaacac agtataaact tcaacttaaa
 4200
 30 agccaggaga ttcacagtct gcagctgaaa gcatgctttg ttcatacagg tgtttataac
 4260
 cttggaactc ctagggtatt tgccaagtta tcggaccaag ttacagtgtt tgaacaagt
 4320
 35 cagcagaatt ccatgcctgc cctgatcctc atcagtaatg tgtgacaact tggaaatttg
 4380
 tactgaaatc cacaataatc agtttttgcg ggatgggttt tacagcagta ttgatatac
 4440
 40 ctaacttggt atggagggtg altgatctct gatccctgca aaatactttg acttgtcatt
 4500
 ttgttgatga tgcaaagcac gttggactga gaatacttaa cattcttttt ctgtatttct
 4560
 45 ttaaaccctg agaataattt acatgctcat aatacaggat atcagcatat ttgtgcacct
 4620
 tattaagccc catcttaaga aaacacaaag tctaagtctg ctgttacaac ttgtcaatgg
 4680
 50 tatacgaata ttaggagatg attctgagaa aggaaaggcc ttgttggcag tactcctggt
 4740
 aagccattag tctctaaatt ccagctttac tgtgaagtcc tatagagtgt taaatacaaa
 4800
 55 ttttctgtc ttgcttcaca cagttcctta aaatcagttt tgaactttgg tcatagagtc
 4860

UcCaLAtLlC agtattttggt ggtccctatg acttatacat aactttgt
4908

5 <210> 17
 <211> 435
 <212> DNA
 <213> human

10 <220>
 <221> misc_feature
 <222> (30)..(30)
 <223> any kind of base

15 <220>
 <221> misc_feature
 <222> (49)..(49)
 <223> any kind of base

20 <220>
 <221> misc_feature
 <222> (75)..(75)
 <223> any kind of base

25 <220>
 <221> misc_feature
 <222> (76)..(76)
 <223> any kind of base

30 <220>
 <221> misc_feature
 <222> (78)..(78)
 <223> any kind of base

35 <220>
 <221> misc_feature
 <222> (79)..(79)
 <223> any kind of base

40 <220>
 <221> misc_feature
 <222> (109)..(109)
 <223> any kind of base

45 <220>
 <221> misc_feature
 <222> (136)..(136)
 <223> any kind of base

50 <220>
 <221> misc_feature
 <222> (137)..(137)
 <223> any kind of base

55 <220>
 <221> misc_feature
 <222> (149)..(149)

<223> any kind of base

 5 <220>
 <221> misc_feature
 <222> (227)..(227)
 <223> any kind of base

 10 <220>
 <221> misc_feature
 <222> (236)..(236)
 <223> any kind of base

 15 <220>
 <221> misc_feature
 <222> (246)..(246)
 <223> any kind of base

 20 <220>
 <221> misc_feature
 <222> (342)..(342)
 <223> any kind of base

 25 <220>
 <221> misc_feature
 <222> (363)..(363)
 <223> any kind of base

 30 <220>
 <221> misc_feature
 <222> (389)..(389)
 <223> any kind of base

 35 <220>
 <221> misc_feature
 <222> (426)..(426)
 <223> any kind of base

 40 <400> 17
 ggtagaaatg attgtgatgt acaaattttn tattttgatc atacttaana agacagagca
 60
 gactcacatt cattnnncna atagtatcac tgtacacata gcgaatttnt ggcgctttta
 120
 45 gattgctctg aaaatnnctg aagagttgnc catagcagcc tggtaagcct tttcctttcc
 180
 cccaaagctc tcctgccctt tgcagaaaga ctgttggtga caactgntgc taactnaata
 240
 50 gcatngggtt gaacttcgcc aaaatecttc cacctcclec catagggcaa caggggtgac
 300
 ttgggcttaa agggcattga gtaagcaagt aggttatcag anaacagagg gaagattcca
 360
 55 ttntagataa tttccaaata ttacaattng tggaactcag agttcaactg ctcagttcct
 420

tcttcngctg accct
435

5

<210> 18
<211> 2224
<212> DNA
<213> human

10

<400> 18
ctttagatct gtgcagcctt tgcgtgccaa acttgtgaaa ttccttttac cttttttgga
60

gtacttgcta taaagccacc tgtaacaaa cccccattat gtacagaata ggacctatcc
120

15

agtagccagg ccagtaggca gttggggaag gtgggaagga tccagcgagg cccctgagcc
180

tgcacctgga cagggtgtacg tctgcaccca tcacctcag caccaggcca cctgcagtc
240

20

cacttactgt actgtgttgt ggaaggatat gctaagtgat gaaagttgag agcagtctca
300

ctggctgtgt aaactttttt ttttttttgg aaattgaagc tgtagagtgc tgcccgaaat
360

25

ctctaggaag ttggtggcaa gggacagcac tcacactctt ctggtcatga tctctgatct
420

ccacctcaaa tgacaataaa aaactgggtcc aacgaagaca ctgctcagca cttcagccat
480

30

caggactaat ccctcgatga ctggaaaaga ggctagcttt gagggaaaaca gcctgggctc
540

ttgggagcag agtccagtgg gtgtgagget gacttgccga cggtcggcag gtaatggctc
600

35

tcagccggcg aggcgggtccc acagctctcc tcccaggga gcctgaggag gaggaggccg
660

ggtgcctgtt tgggtggcagc ttcagcctag ggatacctga agctgttgag caacaccttt
720

40

atgaaatgtt gccagagcag caacacttcc ctgtgggcac agccccggga aatccggtac
780

caagttagca aggtggcagg acccacccaa gcctgatacg catctgggcc cgcggggctc
840

45

agcaggggag gctgctacgg ctgcccactt cccagcaccg tctgtcaggc ttgaaccct
900

ctgtgctgtt cccttcctgg ctaataggga gaccttcgc aggcaccac tgtttcaact
960

50

tgacctccc acccctgct actctctccc acacaccct ccgttcgct agcctacct
1020

gtcagccttt caataaaagt tatgcacaaa tgtgaacacc tgagatggag ctgaacattt
1080

55

cttcactttg ttctttttct gaagtcaaac tcttatcaaa tgccctaaaa ttattaccac

1140

ccaagagaaa caggaaaaag gttacatggt tttgtttact gagagtaaga tcacctgcat
1200

5 ctggaagacg ggctggtaaa ttggtttggc tacagaacag aaagaaaaca aaaacaaacc
1260

tcgtaaggga agtatcgac tcagacacca ccacttccta gagccaaatg agcaatccca
1320

10 aactgcaagt gccgtaagtg ggctgtgac gtcacaccgc ccggcccgag gtatcgcatg
1380

tgcgggggag gccacacta cagctgtcct ctggtctaga aggcaccacc tcgctttcat
1440

15 gtcccggtg ttttgaaaa gcagtatggt gtgtcatgtc tagcggcgaa cacttcctc
1500

cctctgtcct tgaggttgta atataaaaac tgtgtttctg tacgtgtggg tgggaattct
1560

20 ctgacgggtg tcgttcatag cacaagctta cgctgagttc lgaactgtcg ttcacagctg
1620

cggtgtgca tgggtgtgca tctgtgtac ctttggggaa aatttgtatg taaatgtaca
1680

25 gaaataaaaa cgttgcccc aataacagatt tcctctggaa tgtcttcctt acctcacctg
1740

atggtatcca ccgaaggga tttcactacc attaatggtg agtaataaaa tcctccgtgt
1800

30 tcattcagac ctactgcgt cactactttg aacgcctctg taagctgtgt cttcaccgc
1860

cccgaggtg gtggaggag gccctcact ctgcttcgag tcctggctt aaaggtagtc
1920

35 agaggcagag gctggattaa acacacactg tttaccaagt gccactctca gaccacctga
1980

gagacggggg gccatcagta aaattaagag gaattttttt cccttggtcg tgtatgttct
2040

40 gctgatccgt ggctgaagg ttcttagaga cgtcaagaaa tgaatatctt aactgtgat
2100

tctgtgagga aagactggta acccaaaact ctcttctcta atgtattttt taacgaaaat
2160

45 gacaatattt ctttaataaa gtattttatac caaaaaaaaa aaaaaaaaaa aaaaaaaaaa
2220

aaaa
2224

50

<210> 19
<211> 2244
<212> DNA
<213> human

55

<400> 19

EP 1 355 149 A2

gttgtttaaa agcaaggcat gcttgtggat gactctgtaa cagactaatt ggaattgttg
60

5 aagctgctcc ctggttccac tctggagagt aatctgggac atcttagtgt tttgttttgt
120

tttttccct cctctttttt tgggggggag tgtgtgtggg gtttqttttt tagtcttgtt
180

10 tttttaattc attaaccagt ggtagcctt aaggggagga ygacggattg attccacatt
240

ccacttccta gatctagtll agaaaacatg tccccatct ggtgctctta ggaaggagta
300

15 tagtaaatgc ctcatttaat aacatactcc ttttgaaag ttgccttttc tctccacct
360

tgagtagatc cagtatttga tgaaactcat gaaagtgggt ggagcccatc tccccctcc
420

20 tcttttctag gacgcactat atgtgactgt gactttaagg acatttgttt gccatttgtt
480

gatttttttg ggaagttaat ttctaacttc tttcactgat aaatgaagaa aagtattgca
540

25 cctttgaaat gcaccaaag aattgagttt gtaattaaaa aaattttttt tccctttcag
600

tcattgtctt atatgcttag catagatttg cagctcagta gtatatgtgt tcctagaatg
660

30 cagctgaaga cctgttatgt agaggaaata cgaggggtgg tgctagaaga cagacatctg
720

tggaatgatt cacatcctct caagtttaga ggatggaggc ctgcttcatt aagaagctgg
780

35 gggtaggggtg ggggtgggga gaacacttaa caacatgggg accagtcagg ggaatccct
840

tatttctgtt ttgcatatga ggaaccctag agcagccagg tgaggctctc tagtttaata
900

40 aaaatcatgg aaagactctt aatgcagact cttcttaagt gtaaataggg attttttcag
960

cttatttttg ttgcagtttc caatttttaa aaatgttgag gtaatcttcc ccacctccc
1020

45 aaacctaatt cttgtagatg cattagtgtt gaaccaatgc ttctcatgtc tcaatcttgt
1080

atatcatctt ttcagatgta ttaacaaaca aaaccttaaa aagagtagat gaattgccaa
1140

50 acacaattcc taccaataat aaatcgatca actctatcta ttcaggaaag caggaagcat
1200

ttggaaccaca gtgcatgaaa acttcaacat tctgttatta gataatgaat caaccaaag
1260

55 aacaatccag agaaaagaaa attgcaataa taaaaggtaa attaacagaa agataatata
1320

EP 1 355 149 A2

agcaagatag taatagttga ccattctgaa aagcttataa catcactcat catccagcat
1380

5 cctttctgaa aacaaaggat ttttaaata ctttatgcac atatacaaca taggaggttg
1440

gcaaaataat gcactatttc ttaacagcca tgtctcttgt agaacttcaa gttaatctac
1500

10 aaatgaccat tgtgtcttaa tttagattat gaalaccaca ttagtcaggt atttgacta
1560

acccttaata gtatatacag tttctatgga aaattcagtg gtccaaaaat ttccgtagaa
1620

15 tttagagagga cgttggtggg ctgaagatag ctcttgagg gtcactgatg taggtgcaa
1680

tgggggttca caaggccctg acaccgtatt tatagtctaa cctttttatg aaaatctgac
1740

20 tacagctatt taaggagtag tcttaatagc tgaaaatgaa gatagagaaa gacaccaaga
1800

atatgacaca gtttacattc tagtgaggga cacaacaaaa tcaaatttaa aaaagagtgt
1860

25 aatagatgct gataaatact gtagataaag cacataagaa aatagaaata aaggctgtca
1920

atggagaagt catgattttt attttattta tttatttatt tatttgagac agagtcaggc
1980

30 tctgtgcagg ctggagtgca atgggtgtgat ctcgctcact acaacctctg ctctggctc
2040

aagctatcct cccacctcag ctctcaagta gctgggatca cagggtgcgtg ctaccatgcc
2100

35 cggctaattt tttgtagaga tgaggttttg ccatgttgcc caggctggtc tcgaactcct
2160

ggactcaact gacccccct cggcctctca aagtgtgag attataggcg tgcagccggc
2220

40 agctggccat tgtttatggt ctgc
2244

45 <210> 20
<211> / 351
<212> DNA
<213> human

<220>
<221> misc_feature
<222> (62)..(62)
<223> any kind of base

50 <220>
<221> misc_feature
<222> (121)..(121)
<223> any kind of base

55

<220>
 <221> misc feature
 <222> (207)..(207)
 <223> any kind of base
 5

<220>
 <221> misc feature
 <222> (220)..(220)
 <223> any kind of base
 10

<220>
 <221> misc feature
 <222> (276)..(276)
 <223> any kind of base
 15

<220>
 <221> misc feature
 <222> (300)..(300)
 <223> any kind of base
 20

<220>
 <221> misc feature
 <222> (315)..(315)
 <223> any kind of base
 25

<220>
 <221> misc feature
 <222> (336)..(336)
 <223> any kind of base
 30

<400> 20
 tctacttcca catcggcgag accgagaagc gctgtttcat cgaggaaatc cccgacgaga
 60

<400> 120
 cnatggtcat cggcaactat cgtacccaga tgtgggataa gcagaaggag gtcttcctgc
 120

<400> 180
 nctcgacccc tggcctgggc atgcacgtgg aagtgaagga ccccgacggc aagggtggtgc
 180

<400> 240
 tgtcctggca gtacggctcg gagggcnctt tcacgttcan ctcccacacg cccggtgacc
 240

<400> 300
 atcaaactcg tctgcactcc aattcttacc aggatngctc tctttcgctg gtgggcaaan
 300

<400> 351
 tgcgtgttgc atctngacat ccagggttgg gggagnatgc caacaaatta c
 351

<210> 21
 <211> 2631
 <212> DNA
 <213> human
 50

<400> 21
 accttccaac ccagccctcg gctgagccgc gccgcacat gccgcgctg gacaagctcc
 60

<400> 55
 tgctagagga ggcgttgca gacagcccc agactcgctc tttactgagc gtgtttgaag

120

aagatgctgg caccctcaca gactatacca accagctgct ccaggcaatg cagcgcgtct
180

5

atggagccca gaalyagatg tgccctggcca cacaacagct ttctaagcaa ctgctggcat
240

atgaaaaaca gaactttgct cttggcaaag gtgatgaaga agtaatttca acactccact
300

10

atttttccaa agtgggtggat gagcttaatc ttctccatac agagctggct aaacagttgg
360

cagacacaat ggttctacct atcatacaat tccgagaaaa ggatctcaca gaagtaagca
420

15

ctttaaagga tctatttgga ctgcctagca atgagcatga cctctcaatg gcaaaataca
480

gcaggctgcc taagaaaaag gagaatgaga aggtgaagac cgaagtcgga aaagagggtg
540

20

ccgcggcccg gcggaagcag cacctctcct cccltcagta ctactgtgcc ctcaacgcgc
600

tgcagtacag aaagcaaatg gccatgatgg agcccatgat aggctttgcc catggacaga
660

25

ttaacttttt taagaaggga gcagagatgt tttccaaacg tatggacagc tttttatcct
720

ccgttgccaga catggttcaa agcattcagg tagaactgga accgaggcgg aaaagatgcg
780

30

ggtgtcccag caagaattac tttctgttga tgaatctgtt tacactccag actctgatgt
840

ggccgcacca cagatcaaca ggaacctcat ccagaaggct ggttacctta atcttagaaa
900

35

caaaacaggg ctggtcaccg ccacctggga gaggctttat ttcttcaccc aaggcgggaa
960

tctcatgtgt cagcccaggg gagccgtggc tggaggtttg atccaggacc tggacaactg
1020

40

ctcagtgatg gccgtggatt gcgaagaccg gcgctactgc tttcagatca ccacgcccaa
1080

tggaaaatcg ggaataatcc tccaggctga gagcagaaag gaaaatgaag agtggatatg
1140

45

tgcaataaac aacatctcca gacagatcta cctgaccgac aaccctgagg cagtcgcgat
1200

caagttgaat cagaccgctc tgcaagcagt gactcctatt acaagttttg gaaaaaaca
1260

50

agaaagctca tgccccagcc agaacctgaa aaattcagag atggaaaatg aaaatgacaa
1320

gattgttccc aaagcaacag ccagtctacc tgaagcagag gagctgatcg cgcctggagc
1380

55

gccgattcaa ttcgatattg tgcttctctgc tacagaattc cttgatcaga acagagggag

EP 1 355 149 A2

1440

caggcgtacc aacccttttg gtgaaactga ggatgaatca tttccagaag cagaagattc
1500

5

tcttttgcag cagatgttta tagttcgggtt tttgggatca atggcagtta aaacagacag
1560

cactactgaa gtgatttatg aagcgatgag acaagtattg gctgctcggg ctattcataa
1620

10

catcttccgc atgacagaat cccatctgat ggtcaccagc caatctttga gggtgataga
1680

tccacagact caagtatcaa gggccaattt tgaacttacc agtgtcacac aatttgctgc
1740

15

tcatcaagaa aacaagagac tgggttggtt tgtcatccgt gttcctgaat ccactggaga
1800

agaatctctg agtacatata tttttgaaag caactcagaa ggcgaaaaga tatgttatgc
1860

20

tattaatttg ggaaaagaaa ttattgaggt tcagaaggat ccagaagcac tggctcaatt
1920

aatgctgtcc ataccactaa ccaatgatgg aaaatatgta ctgttaaacc atcaaccaga
1980

25

tgacgatgat ggaaatccaa atgaacatag aggcgcagaa tccgaagcat aactcacttg
2040

cgctgtggg ggaagagcga acaggaagga gagctacctc ctaagggtt taacgtctct
2100

30

gacatacagg cacactgacc tgatttccga aggctgacaa tcgtttgtgg aatgtaattc
2160

tgatgccttg atactgagac ttgggagggg aactaagaaa tggttgacag cgttcccacc
2220

35

catctacaat gttatttttag gtgctttgtg gtaagtcttt tttcttagat tgcgctaaaa
2280

tttcttagat tgttcagcgc tcagaacaaa agtttgaaaa atgcattggt catatgaatg
2340

40

tcattctttt tcagtttcca gtatcctttt taaaaaatgg caaaagccta gatttacaat
2400

ttgatgaaca ctaaataattt cttattaata taatctattt ttgtatttta cttaatgagc
2460

45

tttaagtgcc tgcgttctg aaaatttgtt atttataatt cagcttatct cataattgga
2520

cctaatagca tttctttgtg cagttagggt atgagcactg ctttgaggcc caagcactag
2580

50

tagagatgcg cgatacaggt ctagtctcgg taactgttcc agacatcaag c
2631

<210> 22

<211> 2851

<212> DNA

55

EP 1 355 149 A2

<213> human

<400> 22

5 agcatctcag gccatcatcc tgaaacttgg cagccttcgt ggagtataag gacagcatta
60

ttagccatca ttgggtttac tgccaacaaa aggagagggga gccatagggt ctctagatta
120

10 cactcctgag gaaagaagag cacttgccaa aaaatcacao gatttctgtt gtgaaggatg
180

tggtcttgcc atgaaggatg tcctgttgcc tttaaaatct ggaagcgatt caagccaagc
240

15 tgaccaagaa gccaaagaac tggctaggca aataagcttt aaggcagaag tcaattcatc
300

tggaagact atctctgagt cagacttaaa ccactotttt tcaactaactg atttacaaga
360

20 tgatatacct acaacattcc aggggtgctac ggccagtaca tcgtacggac tccagaattc
420

ctcagcagca tcctttcatc aacctacca acctgtagct aagaatacct ccatgagccc
480

25 tcgacagcgc cgggccccagc agcagagtca gagaagggtt tctacttcac cagatgtaat
540

ccagggccac cagccaagag acaaccacac tgatcatggt gggtcagctg tactgattgt
600

30 catcctgact ttggcattgg cagctcttat attccgacga atatatctgg caaacgaata
660

catatttgac tttgagttat aatatggttt tgtgacttat gagctgtgac tcaactgctt
720

35 cattaacat tctgcattgg gtataatcta agaattgttt acaaaaagat tattttgtat
780

tlacccttca ttcctttttt tgatccttgt aagtttagta taaatatatc tagacattca
840

40 gactgtgtct agcagttacg tcctgcttaa agggactaga agtcaaagtt ccttglctca
900

ctatttgatc tgctttgcag ggaaataact tgttttttct catgtttcat cttcttttta
960

45 tgtaaatgtg taatactttc ctatattgcc ctttgaaatt tttggataaa agatgatgtt
1020

ttaagttcca atgagtatta ctagttactc aataccactt attgagtact ctgtttctac
1080

50 gtatgtagaa tgtatagga tagaagagtt gaaaaggga agcaaaactt cttagtggc
1140

ttccttaaaa tgcattcatc aggagatgta ctggaattgc tcattctgtg actttatttg
1200

55 tgtcctaaac attcttcagt gaaaataatt ttatttcagt caaacattta tgaggaaatg
1260

EP 1 355 149 A2

agatcacatc tttgtcactg gatgctactt gaagagggag tactttgtaa ccactttgat
 1320
 5 atgctgttat caccaccccc tgccctctgc tgccataatc acacaaattt aaaaagaaag
 1380
 aaaacagtct tccatagatt ttttaaggaag aaaggqccca agtcaggaga tgcgttygtt
 1440
 10 ttcttcagaa agttaaatgg ggggatctga agatttgaat gttcggctctg ctttgaatg
 1500
 tatgtctttt gggaaatgat tatatgccta gctttataat caggtataaa attttaatta
 1560
 15 ttcccaggaa tatgcataat attgaatatt tcatgtccta ttttaataga aaacctcagg
 1620
 gcccaagtaa ccagtgatag aagttagaaa aacccttta cttagaattg tccacctagt
 1680
 20 cagagcccaa gaaagaattt tcagtygaaa aatcaatata taacttagtg ctagctagcg
 1740
 ccacagactc tagtagataa tattatcatc ataatggctg gtgaaacct ataatcacag
 1800
 25 aaaaacattg ccttcagcat gttcagttcg cagcactgag ggcactcttg aggggtgttg
 1860
 taatgaagat ttaattttta aatacaggtg gttccaagct ttcaaataag ttatgtccca
 1920
 30 aaagtgttat ttgtaagtta atttttttac aagtcaaaca atgttggaag tgggtatttag
 1980
 gttctagatc ggtccacgaa agttagccca tatgtatata ttgaatagta taggggaggg
 2040
 35 tattcataaa gtccttatgt ggttttaact aagtgaatt atggacaaga gaaataattg
 2100
 taaaatcgtc ttaaaggcaa atttaatttt taccctgtt tatgggacat tcttctatt
 2160
 40 aactgtcaga cacaalltct gttttcatct gagagccagg ttccctttat ttctacatct
 2220
 aaaataagaa catattgtac actattatat aatacagaat tgtcttacac ttttaataat
 2280
 45 tgcattttta aagggtgtta caggattatt ttttatatct gtagctgaat ttgttaaagt
 2340
 ctaaaaagct caaggacttt atgaagatct cattatatga ggaaaatcat aggttaccat
 2400
 50 ttataactc tattgccata agaaaataca ctctaaaatc ttgatttgaa acatattaga
 2460
 aaccttgatt cagtgtcag tggctccta gtaagaagtc accgacggtg gcgtcatatg
 2520
 55 agaagaaaga aatccccacc acctcaacct ctgctgagat tgtgtgctag gaacagcctt
 2580

EP 1 355 149 A2

ccctccgttt cccctcagtc aaacttgagc cagcctctgg atcgatgtga tcttattgca
2640

5 tgtttccatg ggggtgtacct atactttaag ccaatcctgc tgcattcact gctaagttaa
2700

ataaaaagcc aagaagaaaa aaaaaatttt gcactgtgca gatcctttgc tatctgactt
2760

10 gcactctcttc cccacactgt cagctagcca cctgcttggt tgtgttgga tatttttttag
2820

cacctgaagc accatctgaa aggggcacca t
2851

15 <210> 23
<211> 3473
<212> DNA
<213> human

20 <400> 23
aagagcagcg gcgagggcgc ggtgggtggct gagtccgtgg tggcagaggc gaaggcgaca
60

gctctagggg ttggcaccgg ccccgagagg aggatgcggg tccggatagg gctgacgctg
120

25 ctgctgtgtg cgggtgctgt gagcttggcc tcggcgctct cggatgaaga aggcagccag
180

gatgaatcct tagattccaa gactactttg acalcagatg agtcagtaaa ggaccacact
240

30 actgcaggca gagtagttgc tggtaaaata tttcttgatt cagaagaatc tgaattagaa
300

tcctctattc aagaagagga agacagcctc aagagccaag agggggagag tgtcacagaa
360

35 gatatcagct ttctagagtc tccaaatcca gaaaacaagg aelatgaaga gccaaagaaa
420

gtacggaaac cagctttgac cgccattgaa ggcacagcac atggggagcc ctgccacttc
480

40 ccttttcttt tcctagataa ggagtatgat gaatgtacat cagatgggag ggaagatggc
540

agactgtggt gtgctacaac ctatgactac aaagcagatg aaaagtgggg cttttgtgaa
600

45 actgaagaag aggtctgctaa gagacggcag atgcaggaag cagaaatggt gtatcaaact
660

ggaatgaaaa tccttaatgg aagcaataag aaaagccaaa aaagagaagc atatcggtat
720

50 ctccaaaagg cagcaagcat gaaccatacc aaagccctgg agagagtgtc atatgctctt
780

ttatttggtg attacttgc acagaatatc caggcagcga gagagatgtt tgagaagctg
840

55 actgaggaag gctctcccaa gggacagact gctcttggtt ttctgtatgc clctggactt
900

EP 1 355 149 A2

ggtgttaatt caagtcaggc aaaggctctt gtatattata catttgagc tcttgygggc
 960
 5 aatctaataag cccacatggt tttgggttac agatactggg ctggcatcgg cgtcctccag
 1020
 agttgtgaat ctgccctgac tcaactatcg cttgttgcca atcatgttgc tagtgatata
 1080
 10 tcgctaacag gaggctcagt agtacagaga atacggctgc ctgatgaagt ggaaaatcca
 1140
 ggaatgaaca gtggaatgct agaagaagat ttgattcaat attaccagtt cctagctgaa
 1200
 15 aaagggtgatg tacaagcaca ggttggctct ggacaactgc acctgcacgg agggcgtgga
 1260
 gtagaacaga atcatcagag agcatctgac tacttcaatt tagcagcaaa tgctggcaat
 1320
 20 tcacatgcca tggccttttt gggaaagatg tattcggaag gaagtgcacat tgtacctcag
 1380
 agtaatgaga cagctctcca ctactttaag aaagctgctg acatgggcaa ccagttgga
 1440
 25 cagagtgggc ttggaatggc ctacctctal gggagaggag ttcaagttaa ttatgatcta
 1500
 gcccttaagt atttccagaa agctgctgaa caaggctggg tggatgggca gctacagctt
 1560
 30 ggttccatgt actataatgg cattggagtc aagagagatt ataaacaggc cttgaagtat
 1620
 tttaatctag cttctcaggg aggccatata ttggctttct ataacctagc tcagatgcac
 1680
 35 gccagtggca cggcgctgat gcgatcatgt cactctgcag tggagttggt taagaatgta
 1740
 tgtgaacgag gccgttggtc tgaaaggctt atgactgcct ataacagcta taaagatggc
 1800
 40 gattacaatg ctgcagtgat ccagtacctc ctcttggtg aacagggcta tgaagtggca
 1860
 caaagcaatg cagcctttat tcttgatcag agagaagcaa gcattgtagg tgagaatgaa
 1920
 45 acttatccca gagctttgct acattggaac agggccgcct ctcaaggcta tactgtggct
 1980
 agaattaagc tcggagacta ccatttctat gggtttgga ccgatgtaga ttatgaaact
 2040
 50 gcatttatcc attaccgtct ggcttctgag cagcaacaca gtgcacaagc tatgtttaat
 2100
 ctgggatata tgcattgagaa aggactgggc attaaacagg atattcacct tgcgaaacgt
 2160
 55 ttttatgaca tggcagctga agccagccca gatgcacaag ttccagtctt cctagccctc
 2220

tgcaaattgg gcgtcgtcta tttcttgag tacatacggg aaacaaacat tcgagatatg
 2280
 5 ttcacccaac ttgatatgga ccagcttttg ggacctgagt gggaccttta cctcatgacc
 2340
 atcattgcgc tgctgttggg aacagtcata gcttacaggc aaaggcagca ccaagacatg
 2400
 10 cctgcacca ggcctccagg gccacggcca gctccacccc agcaggaggg gccaccagag
 2460
 cagcagccac cacagtaata ggcactgggt ccagccttga tcagtgcag cgaaggaagt
 2520
 15 tatctgctgg gaacacttgc atttgattta ggaccttga tcagtggta cctcccagaa
 2580
 gaggcacggc acaaggaagc attgaattcc taaagctgct tagaatctga tgcctttatt
 2640
 20 ttcagggata agtaactctt acctaaactg agctgaatgt ttgtttcagt gccatatgga
 2700
 ataacaactt tcagtggctt tttttttct tttctggaaa catatgtgag acactcagag
 2760
 25 taatgtctac tgtatccagc tatctttctt ggatcctttt ggtcattatt tcagtgtgca
 2820
 taagttctta atgtcaacca tctttaaggt attgtgcac gacactaaaa actgatcagt
 2880
 30 gtaaaaagga aaaccagtt gcaagtttaa acgtgttcca aagtctgaaa atagaacttg
 2940
 ccttttaagt taaaaaaaaa aaaagctatc ttgaaaatgt tttggaactg cgataactga
 3000
 35 gaaactctta ccagtcaca tgcaattaga catattcagc atatttgta ttttaaaagg
 3060
 gaggggtggg aggtttctta ttggtgattg tcacacggta taccatactc ctctccttca
 3120
 40 aagaatgaaa ggccttgta aggagttttt tgtgagcttt acttcttttg aatggaatat
 3180
 acttatgcaa aaccttgta actgactcct tgcactaacg cgagtttgcc ccactactc
 3240
 45 tgtaatttgc ttgtttgttt tgaatataca gacacttgat ccagaagcca gaggatggac
 3300
 taagtgggag aaattagaaa acaaacgaa ctctggttg ggtactacga tcacagacac
 3360
 50 agacatactt ttcctaaagt tgaagcattt gttcccagga tttattttac tttgcatttc
 3420
 cttttgcaca aagaacacat caccatttcc ttttgacaa agaacacac acc
 3473
 55 <210> 24

<211> 401
 <212> DNA
 <213> human

5 <220>
 <221> misc_feature
 <222> (252)..(252)
 <223> any kind of base

10 <220>
 <221> misc_feature
 <222> (303)..(303)
 <223> any kind of base

15 <220>
 <221> misc_feature
 <222> (390)..(390)
 <223> any kind of base

20 <400> 24
 ttagattatt ttcaalltat tattcagaat aaatatatct tttttcttta acttctcaaa
 60
 tagttattga attgtattgg tttaaattaa atgcgtcatg tgtatatatc agtattaatt
 120
 25 caagagatac aaaaggaaat tgagtgaata ataagtctgc ctccctccca tcactctcat
 180
 gtctctacct agaggcaatt attgtcaaca gtttttgatg tgtctttcaa aaaatagtcc
 240
 30 attaagcctg gngtactaga tctcttttaa aagtttaca cctgttacag aatatatata
 300
 aangttcaat tactagtaac accttattac agatacagat tacaacttag gaaatatatt
 360
 35 ttcattggacc attgatgtca ttgggattcn cccctacaat c
 401

40 <210> 25
 <211> 1820
 <212> DNA
 <213> human

45 <400> 25
 aatgtcttag aaaaaggctt tctaaaagaa aaagagcaag aggccatttc ttttcaagat
 60
 agatacaaag aacttcagga aaaacataaa caagaattgg aagacatgag gaaagctggt
 120
 cacgaagccc tcagcattat tgtggatgaa tataaggcac tactgcagtc ttcagttaag
 180
 50 caacaagtag aagctattga aaaacagtac atttctgcaa ttgagaaaca ggcacacaag
 240
 tgtgaggagt tgctaaatgc tcagcatcag aggctccttg aaatgctaga tacagagaag
 300
 55

EP 1 355 149 A2

gaactgttaa aagaaaaaat aaaggaagct ttgattcagc aatctcaaga acagaaggaa
360

5 atattggaaa aqtgttttga ggaagaaagg caaagaaata aagaggcatt agtatccgct
420

gcaaagcttg aaaaagaagc agtgaaggat gcagttttta aagtcgtaga agaagaaaga
480

10 aaaaatttag aaaaagcgca tgctgaagaa agggaattat ggaagacaga acatgcaaaa
540

gatcaagaaa aagtatctca ggaaattcaa aaagctatac aagaacaaag aaaaaataagt
600

15 caggaaactg ltaaggcagc aataatagaa gagcagaaac gaagtgaaaa ggctgtggaa
660

gaggcagtga aaagaacaag agatgaattg atagagtata taaaagaaca gaaaaggctc
720

20 gatcaagtca tccgccaaag aagcctgtcc agtttggaa cgttctctc ctgtgcacag
780

aaacagttaa gtgctttaat agctacggaa ccagttgaca ttgaataaaa agaacatgac
840

25 aaaccacac tggcattgga taaatcatat tacaccttca aaatacacac tctgaattat
900

aaagatgtgt ttgttttctt tccaaatcat gtagaattga tttccagttc aaggataaac
960

30 caaaacaata tttagaacta tcaagtgatc taatttattt tcttttggtt tcttctttac
1020

atttactgtt attttattat tattagtagt agcagcaaca gagtatgata tgacccaaaa
1080

35 gccattgtaa agtgccacat taccaaaatt aattaagtaa actttatagc ctgtgygagt
1140

ctattatata ttattttgca aaagtagtaa atatattatt gtttcatgat gactcttgat
1200

40 gagatgctag aatgtaacca tacatttatc ttattttgag gatagaaata gcatggattt
1260

caacatcact tatltatctg tataattgga aataaaacac cgatatgata gagaatcatt
1320

45 ccggcattac ctaacctctt ctgcagttgg atctatgtat tttcattggt ctactgaaaa
1380

cgaacaatac aattaaaagc actaaagatt attatattaa ttcaactttg atctgatata
1440

50 tcaacttaa c taaaggggtg tgtgtggtgt atgcttgttt cctatttctg ctctttaaag
1500

atactttgaa tcaataaaac cattagtcta caaatcaaat tgtgaactta atctctagaa
1560

55 agagaatata actcagccat ttataggaat ttaggttcaa gtacaggata tatqaaatct
1620

EP 1 355 149 A2

tttcccagta tttcagaatg tacttaattc acaggcagga tgccttcaatg caaaatcatg
 1680
 aatatTTTTa attcaaaaact aaaatgtcat taatatgtat gtatgcaaat gttttatcct
 5 1740
 attttctgaa atgcatctac tttcatgggc tttgtacggt tctgagattt ctcaagtgtaa
 1800
 taaaaagagc tcccaaactt
 10 1820
 <210> 26
 <211> 280
 <212> DNA
 15 <213> human
 <220>
 <221> misc_feature
 <222> (261)..(261)
 <223> any kind of base
 20
 <220>
 <221> misc_feature
 <222> (237)..(237)
 <223> any kind of base
 25
 <400> 26
 tcaagtcata agataaagtt taatcatttg atcatgttaa aagacacaaa acacagccaa
 60
 tctaaccaaa tttcaggcat gcatttacat aaatatatta aattaagaaa agaaattgta
 30 120
 cacttaaacy tccttttcac ctagaatca ttaaattcac agatcaacaa taaaaccaat
 180
 tctctgcatt taccacttca agatacaatt gttctatttt aaagataaca caaactncac
 35 240
 tagtctgggt aggaatttat ntgcattata catatattat
 280
 <210> 27
 <211> 392
 <212> DNA
 40 <213> human
 <400> 27
 45 ttggtttgaa atggcaccac aggactttgg gcctgcctta cttgatagcc tcgttcagtg
 60
 agcaaagact tagtgagcag ctcttgtatg ccaagtattt tgctaagctc tggaaaaaag
 120
 ataaacaaga catggttctt gctttcaagg agtgtgtaat tctttagcca gatatggaaa
 50 180
 cctggaccct gagtgggaga aaggagacag atgaaaggag tccgtgattt tgtaaccaag
 240
 agctgcctgc atggttatga gtatcactga ttttagggac gccacagag ctaaagcatt
 55

300

tttttaatcc gagaagactt ttgtaactca tattagttaa tcttctagct ctgagatagc
360

5

aacacagctc ttagaattct gtaagtaagc tt
392

<210> 28
<211> 2299
<212> DNA
<213> human

<400> 28

10 cgaacccccca cagctggagg gcgaggccag ctgtaccggg cccagtgcc ctttcgaggc
60

15 cacaagcggc cgtcctcctg gtccqgtgct ccggcgccctg atctaggttc atggagccgg
120

20 ggctgtggct ccttttcggg clcacagtga cctccgccgc aggattcgtg ccttgcctcc
180

agtcctgggga tgctggcagg cgcggcgtgt cccaggcccc cactgcagcc agatctgagg
240

25 gggactgtga agagactgtg gctggccctg gcgaggagac tgtggctggc cctggcgagg
300

ggactgtggc cccgacagca ctgcagggtc caagccctgg aagccctggg caggagcagg
360

30 cggccgaggg ggccctlgag caccaccgat ccaggcgctg cacgtgcttc acctacaagg
420

acaaggagtg tgtctactat tgccacctgg acatcatttg gatcaacact cccgaacaga
480

35 cggtgcccta tggactgtcc aactacagag gaagcttcg gggaagagg tctgcggggc
540

cacttcagg gaatctgcag ctctcacalc ggccacactt gcgctgcgct tgtgtgggga
600

40 gatatgacaa ggccctgcctg cacttttgca cccaaactct ggacgtcagc agacaggttg
660

aagtcaagga ccaacaaagc aagcaggctt tagacctcca ccatccaaag ctcatgcccg
720

45 gcagtggact cggcctcgct ccatctacct gcccccgctg cctctttcag gaaggagccc
780

cttaggagga caggcctgca gcacctcgtt ctggggaggc ttctgtcatt gctcacacac
840

50 agttcagatt tccacctctt tatagacaag aagtgaattt gcctggggca gaacacccac
900

ccaaagagtc cccacttaac aatacccccc cccacggca agaattgccca aatccgaatg
960

55 accccagttt tcctaagag taaaatgatc ccagatgtgc cccagagcat gacgcctgca
1020

EP 1 355 149 A2

gctccggttt catgcaggaa attggttttg gagagttttg gcaagttgga aagccactta
1080

5 ctggcctttt acatgacttc tcttgagaa taagtggact ccaagctaac tctttgcaaa
1140

tgtaaacaca tgtccatctt gtaataaatg caaaatgcc gtgcagcaga agcatgcgac
1200

10 tttcatatcc ttgcctagaa taggctgcat ggtgtatgtc agtgagggcc acgaggcgtc
1260

ggcttttagac acagatcata gctctacagg agtttatgaa tttgaagctt atgggatttt
1320

15 ggagagaaa ttttcagctg tgcttgatac ccacaaaag aatgtatctc gaaagaatga
1380

aggaagaaga aaaaaggatc cttgatgttt gtgacaagaa aatgagaaa ttagtlatctg
1440

20 caatacagag cttgttctctg ttcagtgact gacctctctgt attctgtata gacaccaggc
1500

cgatacacag tggagttccc aggccttgtt tgcaggaagc cgactgtaaa gacagcccca
1560

25 gctcaaggct attaggttga atatttgctt tcatgagtaa atgtggatct ttggggaatg
1620

gcttcaaaat aagtcacgaa cacaattctt ttgtaaatta tgtaaattcc tgtttatata
1680

30 aattggcaac aacttatacc gtctgacagt tcaaaatctc tttcagctgc gctcttccca
1740

ccgagccgag cttactgtga gtgtggagat gttatccac catgtaaagt cgctgcgca
1800

35 ggggagggt gcccctctcc ccaaccagc cacagagaga taggaaacgg catttgagtg
1860

ggtgtccagg gccccgtaga gagacattta agatgggtga tgacagagca ttggccttga
1920

40 ccaaagtta aatcctctgt gtgtatttca taagttatta caggataaa agtgatgacc
1980

tatcatgagg aaatgaaagt ggctgatttg ctggtaggat tttgtacagt ttagagaagc
2040

45 gattatttat tgtgaaactg ttctccactc caactcctt atgtggatct gttcaaagta
2100

gtcactgtat atacgtatag agaggtagat aggtaggtag attttaaatt gcattctgaa
2160

50 taaaaactca tactccttag agcttgaatt acatttttaa aatgcatatg tgctgtttgg
2220

caccgtggca agatggatc agagagaaac ccatcaattg ctcaaatact cagaaagtac
2280

55 tgtcaaaagc ctaataaaa
2299

EP 1 355 149 A2

<210> 29
 <211> 1339
 <212> DNA
 <213> human

5
 <400> 29
 ctaaacaaaa tcattcactt ccctgatttt gataagaaaa ttctgtataa gctgtttcct
 60

10
 ctgcctctcc tctacgttgg aaaccacata agtggattat caagcacaag taaattaagc
 120
 ctaccgatgt tcaccgtgct caggaaattc accattccac ttaccttact tctggaaacc
 180

15
 atcatacttg ggaagcagta ttcactcaac atcatcctca gtgtctttgc cattattctc
 240
 ggggctttca tagcagctgg gtctgacctt gcttttaact lagaaggcta tatttttcta
 300

20
 ttctgaatg atatcttcac atcagcaaat ggagtttata ccaaacagaa aatggaccca
 360
 aaggagctag ggaaatcagg agtacttttc tacaatgcct gcttcatgat tatcccaact
 420

25
 cttattatta gtgtctccac tggagacctc caacaggcta ctgaattcaa ccaatggaag
 480
 aatgttgtgt ttatcctaca gtttcttctt tctgtttttt tggggtttct gctgatgtac
 540

30
 tccacggttc tgtgcagcta ttacaattca gccctgacga cagcagtggg tggagccatc
 600
 aagaatgtat ccgttgccta cattgggata ttaatcgggtg gagactacat tttctctttg
 660

35
 ttaaactttg taggggttaa tatttgcatt gcagggggct tgagatattc cttttlaaca
 720
 ctgagcagcc agttaaacc taaacctgtg ggtgaagaaa acatctgttt ggatttgaag
 780

40
 agctaaagag tctgcagcag gattggagac tgacttgtga ctgcgggctg ggggggcatt
 840
 cccagtagga atgtgaagcc agaggtttcg gattcgtgac atccaccccc tgggcaagtg
 900

45
 agagcatctg caaaatgcaa agagaactac ctcatatgca ggatgagcca atggcagttc
 960
 caagaaatgt actcgggcga caccttacct gtggaaagca aatcttttca aaataagcca
 1020

50
 ctgggactcg gtaggtggag cccagctgc tcttctaggg acctatgggg ccttcgtggc
 1080
 atctctgtgc tgtgtgctgg ggaggaggtt gatgtaatgg tgactctttt ctgatcagca
 1140

55
 ccttggccgt gattcccaag gtcccagcca aagcaaaggg ccagttgttt cagtttaaac
 1200

agacatgtct ttagtctaataaaaattagtt aactgccagt aaagttattt gttagctllly
1260

5 atgaaagcta tgttggtatc tttccctaataat catcaaagta aataaaaaat catttctatg
1320

taaaaaaaaa aaaaaaaaaa
1339

10 <210> 30
<211> 4250
<212> DNA
<213> human

15 <400> 30
gaacacatcg cgtttgcatac ccagaaagta gtcgccgcga ctatttcccc caaagagaca
60

agcacacatg taggaatgac aaaggcttgc gaaggagaga gcgcagcccg cggcccggag
120

20 agatcccctc gataatggat tactaaatgg gatacacgct gtaccagttc gctccgagcc
180

ccggccgcct gtccgtcgat gcaccgaaaa ggggtgaagta gagaaataaa gtctccccgc
240

25 tgaactacta tgaggtcaga agccttgctg ctatatttca cactgctaca ctttgcctggg
300

gctggtttcc cagaagattc tgagccaatc agtatttcgc atggcaacta taaaaaacag
360

30 tatccggtgt ttgtgggcca caagccagga cggaacacca cacagaggca caggctggac
420

atccagatga ttatgatcat gaacggaacc ctctacattg ctgctaggga ccatatttat
480

35 actgttgata tagacacatc acacacggaa gaaatttatt gtagcaaaaa actgacatgg
540

aaatctagac aggccgatgt agacacatgc agaataagg gaaaacataa ggatgagtgc
600

40 cacaacttta ttaaagttct tctaaagaaa aacgatgatg cattgtttgt ctgtggaact
660

aatgccttca acccttcctg cagaaactat aagatggata cattggaacc attcggggat
720

45 gaattcagcg gaatggccag atgcccatat gatgccaaac atgccaacgl tgcactgttt
780

gcagatggaa aactatactc agccacagtg actgacttcc ttgccattga cgcagtcatt
840

50 taccggagtc ttggagaaag ccctaccctg cggaccgtca agcacgattc aaaatggttg
900

aaagaacat actttgttca agccgtggat tacggagatt atatctactt cttcttcagg
960

55 gaaatagcag tggagtataa caccatggga aaggtagttt tcccaagagt ggctcaggtt

1020

tgtaagaatg atatgggagg atctcaaaga gtcctggaga aacagtggac gtcgttcctg
1080

5

aaggcgcgct tgaactgctc agttcctgga gactctcatt tttatttcaa cattctccag
1140

gcagttacag atgtgattcg tatcaacggg cgtgatgttg tcctggcaac gttttctaca
1200

10

ccttataaca gcatccctgg gtctgcagtc tgtgcctatg acatgcttga cattgccagt
1260

gtttttactg ggagattcaa ggaacagaag tctcctgatt ccacctggac accagttcct
1320

15

gatgaacgag ttcctaagcc caggccaggt tgctgtgctg gtcctcctc cttagaaaga
1380

tatgcaacct ccaatgagtt ccctgatgat accctgaact tcatcaagac gcacccgctc
1440

20

atggatgagg cagtgcctc catcttcaac aggccatggg tcctgagaac aatggtcaga
1500

taccgcctta ccaaaattgc agtggacaca gctgctgggc catatcagaa tcacactgtg
1560

25

gtttttctgg gatcagagaa ggaatcatc ttgaagttt tggccagaat aggaaatagt
1620

ggttttctaa atgacagcct tttcctggag gagatgagtg tttacaactc tgaaaaatgc
1680

30

agctatgatg gagtcgaaga caaaaggatc atgggcatgc agctggacag agcaagcagc
1740

tctctgtatg ttgcgttctc tacctgtgtg ataaaggttc cccctggccg gtgtgaacga
1800

35

catgggaagt gtaaaaaaac ctgtattgcc tccagagacc catattgttg atggataaag
1860

gaagggtgtg cctgcagcca tttatcacc caccagcagac tgacttttga gcaggacata
1920

40

gagcgtggca atacagatgg tctgggggac tgtcacaatt cctttgtggc actgaatgac
1980

atttcaactc ctctaccaga taatgaaatg tcttacaaca cagtglatgg gcattccagt
2040

45

tcctcttgc ccagcacaac cacatcagat tcgacggctc aagaggggta tgagtctagg
2100

ggaggaatgc tggactggaa gcatctgctt gactcacctg acagcacaga ccctttgggg
2160

50

gcagtgtctt ccataatca ccaagacaag aaggagtgga ttcgggaaag ttacctcaaa
2220

ggccacgacc agctggttcc cgtcacctc ttggccattg cagtcctcct ggctttcgtc
2280

55

atgggggccc tcttctcggg catcacctc tactgcgtct gtgatcatcg gcgcaaagac

EP 1 355 149 A2

2340

gtggctgtgg tgcagcgcaa ggagaaggag ctcacccact cgcgccgggg ctccatgagc
2400

5

agcgtcacca agctcagcgg cctctttggg gacactcaat ccaaagaccc aaagccggag
2460

gccatcclea cgcactcal gcacaacygc aagctcgcca ctcccggcaa cagggccaag
2520

10

atgctcatta aagcagacca gcaccacctg gacctgacgg ccctcccccac ccagaggtca
2580

accccaacgc tgcagcagaa gcggaagccc agccgcggca gccgcgagtg ggagaggaaac
2640

15

cagaacctca tcaatgcctg cacaaggagc atgccccca tgggctcccc tgtgattccc
2700

acggacctgc ccctgcgggc ctccccagc cacatcccca gcgtggtggt cctgcccac
2760

20

acgcagcagg gctaccagca tgagtacgtg gaccagccca aaatgagcga ggtggcccag
2820

atggcgctgg aggaccaggc cgccacactg gagtataaga ccatcaagga acatctcagc
2880

25

agcaagagtc ccaaccatgg ggtgaacctt gtggagaacc tggacagcct gcccccaaa
2940

gttcacagc ggagggcctc cctgggtccc ccgggagcct ccctgtctca gaccggtcta
3000

30

agcaagcggc tggaatatga ccaactcctt tcctacgggg ttgactataa gaggagctac
3060

cccacgaact cgctcacgag aagccaccag gccaccactc taaaagaaa caacactaac
3120

35

tcctccaatt cctctcacct ctccagaaac cagagctttg gcaggggaga caaccgcgcg
3180

cccgccccgc agaggggtgga ctccatccag gtgcacagct ccagccatc tggccaggcc
3240

40

gtgactgtct cgaggcagcc cagcctcaac gcctacaact cactgacaag gtcggggctg
3300

aagcgtacgc cctcgctaaa gccggacgta cccccaaac catcctttgc tcccctttcc
3360

45

acatccatga agcccaatga tgcgtgtaca taatcccagg gggagggggg caggtgtcga
3420

accagcaggc aaggcgaggt gcccgctcag ctacgaagg ttctcaactg cctcgagtac
3480

50

ccaccagacc aagaaggcct gcggcagagc cgaggacgct gggtcctcct ctctgggaca
3540

caggggtact cacgaaaact gggccgcgtg gtttggtgaa ggtttgcaac ggcggggact
3600

55

caccttcatt ctcttccttc actttcccc acacctaca acaggtcgga cccacaaaag

EP 1 355 149 A2

3660

acttcagtta tcatacacaaa catgagccaa aagcacatac ctaccccatc cccacacccc
3720

5

acacacacac acacatgcac acacacacata cacacacacg cacagaggtg aacagaaact
3780

gaaacatttt gtccacaact tcacgggacg tggccagact gggtttgctg tccaacctgc
3840

10

aaaacacaaa tacatttttt aaaatcaaga aaatttaaaa agacaaaaaa aaaagaattc
3900

attgataatt ctaactcaga cttaacaat ggcagaagtt tactatgcgc aaatactgtg
3960

15

aatgcccgc cagtgttaca gctttctgtt gcagcagala aatgccatgt tgggcaacta
4020

tgtcatagat ttctgctcct cctctctttt aatgaaataa cgtgaccgtt aacgcaagta
4080

20

actctttatt tattgttcac cctttttttt cttaaggaaa ggactcttcc aaatatcatc
4140

ctatgaacag ctcttcagaa agcccattga aagttaaact atttaacgtg aaatccatta
4200

25

actggaataa ttgagtttct ttattttttac aataaattca ctgagtaaat
4250

<210> 31
<211> 2785
<212> DNA
<213> human

30

<400> 31
ctttagccca acagtcaaaa ataattgatg ctaccctaca aatgtccaaa actctagtat
60

35

atcatatttc taagttacag caaatattag tcttgctaaa ccagggagct ttggcaaaaa
120

tgttttttga cagtaaattt gtccttgatt atatattaac tagtcaaaga ggtgtttgta
180

40

acattattag agcttcttgc ttaggtggg ttaacaccac caatcaagag gtcattctaa
240

cagaaagcct ggatcagaaa accatcacc ctaaaaaaca tgccttacat atttaacaca
300

45

ctctgaaatc cagtcaaaat atgactaaag gcccttgcca tgactgatgt attctcctgg
360

ccaacgccaa acaaatggga gcctgggtac gagtcagcct tcagggactt gtcacatttc
420

50

tacttggttt cttccttggt attgtcataa taaaatgttt tctatgctgt ttagtgcaac
480

ttaggcccta ttctgtagaa gtctcctcta ctattcaggc cactcaaaca ccccaaataa
540

55

EP 1 355 149 A2

ttgagttcaa aatcgacatc aagatataaa ggaatcagtg actaaatata tttcatalat
 600
 5 ggtattttta ttgattattg tgctgtcttg acctagtatg gaggccttgg cttagaggctg
 660
 gtcagtttcc tctcttgagc agctgattaa atccacaacc caaccacttc ccttatcagg
 720
 10 ttctcacact ctggggccac tatgtaccca ctctaatac cacagggcca gacatcagac
 780
 aattaaggac agcgcccatg ccccaaagcc cgccaaaatt atgcaaatta ttcaaaatta
 840
 15 ttcaacctag ctaacccac cctttttgct gtacataagc tgccattcc cctccagcc
 900
 tgtggtaccc agtcctcagg lgcaaccccc lgcgtggtcc tctgtggcag cttctctca
 960
 20 ttcagagctg tttccacag aggtagtga aagaactgga tttcaagtt cactttgcaa
 1020
 gagaaaaaga aaactcagta gaagataatg gcaagtcag actggggata tgatgacaaa
 1080
 25 aatggtcctg aacaatggag caagctgtat ccatttgcca atggaaataa ccaatccct
 1140
 gttgatatta aaaccagtga aaccaaacat gacaccttc tgaaacctat tagtgtctcc
 1200
 30 tacaaccag ccacagccaa agaaattatc aatgtggggc attctttcca tgtaaatttt
 1260
 gaggacaacg ataaccgatc agtgctgaaa ggtggtcctt tctctgacag ctacaggctc
 1320
 35 tttcagtttc attttcactg gggcagtaca aatgagcatg gttcagaaca tacagtggat
 1380
 ggagtcaaat attctgccga gcttcacgta gctcactgga attctgcaa gtactccagc
 1440
 40 cttgctgaag ctgcctcaaa ggctgatggg ttggcagtta ttggtgttt gatgaagggt
 1500
 ggtgaggcca acccaaagct gcagaaagta cttgatgcc tccaagcaat taaaaccaag
 1560
 45 ggcaaacgag cccattcac aaatttgac ccctctact ccttccttc atccctggat
 1620
 ttctggacct accctggctc tctgactcat cctctcttt atgagagtgt aacttggatc
 1680
 50 atctgtaagg agagcatcag tgtcagctca gagcagctgg cacaattccg cagccttcta
 1740
 tcaaatgttg aaggtgataa cgctgtcccc atgcagcaca acaaccgcc aaccaacct
 1800
 55 ctgaaggcca gaacagtga agcttcattt tgatgattct gagaagaaac ttgtccttcc
 1860

EP 1 355 149 A2

tcaagaacac agccctgctt ctgacataat ccagttaaaa taataatttt taagaaataa
1920

5 atttatttca atattagcaa gacagcatgc cttcaaatca atctgtaaaa ctaagaaact
1980

taaatttttag ttcttactgc ttaattcaaa taataattag taagctagca aatagtaac
2040

10 tgtaagcata agcttatctt aaattcaagt ttagtttgag gaattcttla aaattacaac
2100

taagtgattt gtatgtctat ttttttcagt ttattlgaac caataaaata attttatctc
2160

15 tttctttctg ttgtgcattc agtttctaaa accattaagt ttctactcca tttacattca
2220

aaaatcttaa atactttact tgcaagagta ttttgcttca aatacaacaa cctaagagca
2280

20 gctggagatg aaatattggg aaattcattt gcttactcct gaagacaaaa atatagctga
2340

gatgaccact ggatttaata tcgttatgct ggcccaacat tgctaccatt tgtgtgtct
2400

25 gtgatcaaaa tgattatctt ttatatagga agatgacgct tctggatatt gctttcactt
2460

cttctcccca cgtagcaag gacaatgctt ctctgccatt attacaacta gttagttlgc
2520

30 atggagaatc tttactttaa aattggaaga aaagtcacaa glgaatggtt tataaaaaatg
2580

ctaaagaagt cattcttgct tagaatcata tagaaacatc atgcaatctt ttagtcagat
2640

35 gtgcgcttca ccttatgcta tttttatctt taattgacac acaataattg tacatgttta
2700

tggagtatag tgtggtgttt tctgtttgtt tgtttgtttt ttgagacaag gtctcactct
2760

40 gccagtcagg gtggagtgcg atggt
2785

<210> 32
<211> 9588
<212> DNA
45 <213> human

<400> 32
ccgaccaaca ccaacaccca gctccgacgc agctcctctg cgccttgcc gccctccgag
60

50 ccacagcttt cctcccgtc ctgcccccg cccgtcgcg tctccgcgt cgcagcgcc
120

tcgggagggc ccaggtagcg agcagcgacc tcgcgagcct tccgcactcc cgcccggttc
180

55 cccggcgcgc cgcctatcct tggccccctc cgctttctcc gcgcgggcc gccctcgctta
240

tgccctcggcg ctgagccgct ctcccgattg cccgccgaca tgagctgcaa cggaggctcc
 300
 5 caccgcgga tcaacactct yggccgcatg atccgcgcg agtctggccc ggacctgcgc
 360
 tacgaggtga ccagcggcgg cgggggcacc agcaggatgt actattctcg gcgcggcgtg
 420
 10 atcaccgacc agaactcgga cggctactgt caaacccgga cgatgtccag gcaccagaac
 480
 cagaacacca tccaggagct gctgcagaac tgctccgact gcttgatgcg agcagagctc
 540
 15 atcgtgcagc ctgaattgaa gtatggagat ggaatacaac tgactcggag tcgagaattg
 600
 gatgagtgtt ttgcccaggc caatgaccaa atggaaatcc tcgacagctt gatcagagag
 660
 20 atgcggcaga tgggccagcc ctgtgatgct taccagaaaa ggcttcttca gctccaagag
 720
 caaatgcgag ccctttataa agccatcagt gtccctcgag tccgcagggc cagctccaag
 780
 25 ggtggtggag gctacacttg tcagagtggc tctggctggg atgagttcac naaacatgtc
 840
 accagtgaat gtttggggtg gatgaggcag caaagggcgg agatggacat ggtggcctgg
 900
 30 ggtgtggacc tqgcctcagt ggagcagcac attaacagcc accggggcat ccacaactcc
 960
 atcggcgact atcgtggca gctggacaaa atcaaagccg acctgcgcga gaaatctgcg
 1020
 35 atctaccagt tggaggagga gtatgaaaac ctgctgaaag cgtcctttga gaggatggat
 1080
 cacctgcgac agctgcagaa catcattcag gccacgtcca gggagatcat gtggatcaat
 1140
 40 gactgcgagg aggaggagct gctgtacgac tggagcgaca agaacaccaa catcgctcag
 1200
 aaacaggagg ccttctccat acgcatgagt caactggaag ttaaagaaaa agagctcaat
 1260
 45 aagctgaaac aagaaagtga ccaacttgct ctcaatcagc atccagcttc agacaaaatt
 1320
 gaggcctata tggacactct gcagacgcag tggagttgga ttcttcagat caccaagtgc
 1380
 50 attgatgttc atctgaaaga aaatgctgcc tactttcagt tttttgaaga ggcgcagtct
 1440
 actgaagcat acctgaaggg gctccaggac tccatcagga agaagtacc ctcgcgacaag
 1500
 55 aacatgcccc tgcagcacct gctggaacag atcaaggagc tggagaaaga acgagagaaa
 1560

atccttgaat acaagcglca ggtgcayaac ttggtaaaca agtctaagaa gattgtacag
 1620
 5 ctgaagcctc gtaaccaga ctacagaagc aataaaccca ttattctcag agctctctgt
 1680
 gactacaaac aagatcagaa aatcgtgcat aagggggatg agtgtatcct gaaggacaac
 1740
 10 aacgagcgca gcaagtggta cgtgacgggc ccgggaggcg ttgacatgct tggtccctct
 1800
 gtggggctga tcatccctcc tccgaaccca ctggccgtgg acctctcttg caagattgag
 1860
 15 cagtactacg aagccatctt ggctctgtgg aaccagctct acatcaacat gaagagcctg
 1920
 gtgtcctggc actactgcat gattgacata gagaagatca gggccatgac aatcgccaag
 1980
 20 ctgaaaacaa tgcggcagga agattacatg aagacgatag ccgaccttga gttacattac
 2040
 caagagttca tcagaaatag ccaaggctca gagatgtttg gagatgatga caagcggaaa
 2100
 25 atacagtctc agttcaccga tgcccagaag cattaccaga ccttggtcat tcagctccct
 2160
 ggctatcccc agcaccagac agtgaccaca actgaaatca ctcatcatgg aacctgcca
 2220
 30 gatgtcaacc ataataaagt aattgaaacc aacagagaaa atgacaagca agaaacatgg
 2280
 atgctgatgg agctgcagaa gattcgcagg cagatagagc actgcgaggg caggatgact
 2340
 35 ctcaaaaacc tccctctagc agaccagggg tcttctcacc acatcacagt gaaaattaac
 2400
 gagcttaaga gtgtgcagaa tgattcaca gcaattgctg aggttctcaa ccagcttaaa
 2460
 40 gatatgcttg ccaacttcag aggttctgaa aagtactgct atttacagaa tgaagtattt
 2520
 ggactatttc agaaactgga aaatatcaat ggtgttacag atggctactt aaatagctta
 2580
 45 tgcacagtaa gggcactgct ccaggctatt ctccaaacag aagacatggt aaaggtttat
 2640
 gaagccaggc tcaactgagga ggaaactgtc tgccctggacc tggataaagt ggaagcttac
 2700
 50 cgctgtggac tgaagaaaat aaaaaatgac ttgaacttga agaagtcgtt gttggccact
 2760
 atgaagacag aactacagaa agcccagcag atccactctc agacttcaca gcagtatcca
 2820
 55 ctttatgata tggacttggg caagtccggt gaaaaagtca cacagctgac agaccgctgg
 2880

EP 1 355 149 A2

caaaggatag ataaacagat cgactttaga ttatgggacc tggagaaaca aatcaagcaa
2940

5 ttgaggaatt atcgtgataa ctatcaggct ttctgcaagt ggctctatga tcgtaaacgc
3000

cgccaggatt ccttagaatc catgaaattt ggagattcca acacagtcac gcggtttttg
3060

10 aatgagcaga agaacttgca cagtgaataa tctggcaaac gagacaaatc agaggaaagta
3120

caaaaaattg ctgaactttg cgccaattca attaaggatt atgagctcca gctggcctca
3180

15 tacacctcag gactggaaac tctgctgaac atacctatca agaggaccat gattcagtcc
3240

ccttctgggg tgattctgca agaggctgca gatgttcacg ctcggtacat tgaactactt
3300

20 acaagatctg gagactatta caggttctta agtgagatgc tgaagagttt ggaagatctg
3360

aagctgaaaa ataccaagat cgaagttttg gaagaggagc tcagactggc ccgagatgcc
3420

25 aactcggaag actgtaataa gaacaaattc ctggatcaga acctgcagaa ataccaggca
3480

gagtgttccc agttcaaagc gaagcttgcg agcctggagg agctgaagag acaggctgag
3540

30 ctggatggga agtcggctaa gcaaaatcta gacaagtgc acggccaaat aaaagaactc
3600

aatgagaaga tcacccgact gacttatgag attgaagatg aaaagagaag aagaaaatct
3660

35 gtggaagaca gatttgacca acagaagaat gactatgacc aactgcagaa agcaaggcaa
3720

tgtagaaagg agaaccttgg ttggcagaaa ttagagtctg agaaagccat caaggagaag
3780

40 gagtacgaga ttgaaagggt gagggttcta ctgcaggaag aaggcaccgc gaagagagaa
3840

tatgaaaatg agctggcaaa ggtaagaaac cactataatg aggagatgag taatttaagg
3900

45 aacaagtatg aaacagagat taacattacg aagaccacca tcaaggagat atccatgcaa
3960

aaagaggatg attccaaaaa tcttagaaac cagcttgata gactttcaag ggaaaatcga
4020

50 gatctgaagg atgaaattgt caggctcaat gacagcatct tgcaggccac tgagcagcga
4080

agcgagctg aagaaaacgc ccttcagcaa aaggcctgtg gctctgagat aatgcagaag
4140

55 aagcagcatc tggagataga actgaagcag gtcatgcagc agcgctctga ggacaatgcc
4200

cggcacaagc agtccctgga ggaggctgcc aagaccattc aqgacaaaa taaggagatc
 4260
 5 gagagactca aagctyagtt tcaggaggag gccaaagcgc gctgggaata tgaaaalgaa
 4320
 ctgagtaagg taagaacaa ttatgatgag gagatcatta gcttaaaaa tcagtttgag
 4380
 10 accgagatca acatcaccaa gaccaccatc caccagctca ccatgcagaa ggaagaggat
 4440
 accagtggct accgggctca gatagacaat ctcacccgag aaaacaggag cttatctgaa
 4500
 15 gaaataaaga ggctgaagaa cactctaacc cagaccacag agaattctcag gaggggtgaa
 4560
 gaagacatcc aacagcaaaa ggccactggc tctgaggtgt ctcagaggaa acagcagctg
 4620
 20 gaggttgagc tgagacaagt cactcagatg cgaacagagg agagcgtaag atataagcaa
 4680
 tctcttgatg atgctgcaa aaccatccag gataaaaaca aggagataga aaggttaaaa
 4740
 25 caactgatcg acaagaaac aaatgaccgg aaatgcctgg aagatgaaaa cgcgagatta
 4800
 caaagggtec agtatgacct gcagaaagca aacagtagtg cgacggagac aataaacaaa
 4860
 30 ctgaaggltc aggagcaaga actgacacgc ctgaggatcg actatgaaag ggtttcccag
 4920
 gagaggactg tgaaggacca ggatatcagc cggttccaga actctctgaa agagctgcag
 4980
 35 ctgcagaagc agaagggtgga agaggagctg aatcggtga agaggaccgc gtcagaagac
 5040
 tcctgcaaga ggaagaagct ggaggaagag ctggaaggca tgaggaggtc gctgaaggag
 5100
 40 caagccatca aaatcaccaa cctgaccacg cagctggagc aggcattcat tgtaagaag
 5160
 aggagtgagg atgacctccg gcagcagagg gacgtgctgg atggccacct gagggaaaaag
 5220
 45 cagaggaccc aggaagagct gaggaggctc tcttctgagg tcgaggccct gaggcggcag
 5280
 ttactccagg aacaggaaag tgtcaaaca gctcacttga ggaatgagca tttccagaag
 5340
 50 gcgatagaag ataaaagcag aagcttaaat gaaagcaaaa tagaaattga gaggctgcag
 5400
 tctctcacag agaacctgac caaggagcac ttgatgttag aagaagaact gcggaacctg
 5460
 55 aggtggagt acgatgacct gaggagagga cgaagcgaag cggacagtga taaaaatgca
 5520

accatcttgg aactaaggag ccagctgcag atcagcaaca accggaccct ggaactgcag
 5580

5 gggctgatta atgatttaca gagagagagg gaaaatttga gacaggaaat tgagaaattc
 5640

caaaagcagg ctttagaggc atctaatagg attcaggaat caaagaatca gtgtactcag
 5700

10 gtggtacagg aaagagagag ctttctggtg aaaatcaaag tcctggagca agacaaggca
 5760

aggcctgcaga ggctggagga tgagctgaat cgtgcāaaat caactctaga ggcagaaacc
 5820

15 agggtgaaac agcgccctgga gtgtgagaaa cagcaaattc agaatgacct gaatcagtgg
 5880

aagactcaat attcccgcaa ggaggagcct attaggaaga tagaatcgga aagagaaaag
 5940

20 aqtgagagag agaagaacag tcttaggagt gagatcgaaa gactccaagc agagatcaag
 6000

agaattgaag agagggtgcag gcgtaagctg gaggattcta ccagggagac acagtcacag
 6060

25 ttagaaacag aacgctcccg atatcagagg gagattgata aactcagaca gcgcccatac
 6120

gggctcccatc gagagaccca gactgagtgt gagtggaccg ttgacacctc caagctgggtg
 6180

30 tttgatgggc tgaggaagaa ggtgacagca atgcagctct atgagtgtca gctgatcgac
 6240

aaaacaacct tggacaaact attgaagggg aagaagtcag tggaagaagt tgcttctgaa
 6300

35 atccagccat tccttcgggg tgcaggatct atcgctggag catctgcttc tcctaaggaa
 6360

aaatactctt tggtagaggc caagagaaaag aaattaatca gcccagaatc cacagtcatg
 6420

40 cttctggagg ccagggcagc tacagggtgt ataattgatc cccatcggaā tgagaagctg
 6480

actgtcgaca gtgccatagc tcgggacctc attgacttcg atgaccgtca gcagatatat
 6540

45 gcagcagaaa aagctatcac tggttttgat gatccatttt caggcaagac agtatctggt
 6600

tcagaagcca tcaagaaaaa tttgattgat agagaaaccg gaatgcccct gctggaagcc
 6660

50 cagattgctt cagggggtgt agtagaccct gtgaacagtg tctttttgcc aaaagatgtc
 6720

gccttgcccc gggggctgat tgatagagat ttgtatcgat ccctgaatga tccccgagat
 6780

55 agtcagaaaa actttgtgga tccagtcacc aaaaagaagg tcagttacgt gcagctgaag
 6840

gaacgggtgca gaatcgaacc acatactggg ctgctcttgc tttcagtaca gaagagaagc
 6900

5 atgtccttcc aaggaatcag acaacctgtg accgtcactg agctagtaga ttctgggtata
 6960

ttgagaccgt ccactgtcaa tgaactggaa tctggtcaga tttcttatga cgagggtggg
 7020

10 gagagaatta aggacttcct ccagggttca agctgcatag caggcatata caatgagacc
 7080

acaaaacaga agcttggcat ttatgaggcc atgaaaattg gcttagtccg acctgggtact
 7140

15 gctctggagc tgctggaagc ccaagcagct actggcttla tagtggatcc tgttagcaac
 7200

ttgaggttac cagtggagga agcctacaag agaggtctgg tgggcattga gttcaaagag
 7260

20 aagctcctgt ctgcagaacg agctgtcact gggataatg atcctgaaac aggaaacatc
 7320

atctctttgt tccaagccat gaataaggaa ctcacgaaa agggccacgg tattecgtta
 7380

25 ttagaagcac agatcgcaac cgggggggac attgaccaa aggagagcca tcgtttacca
 7440

gttgacatag catataagag gggctatttc aatgaggaac tcagtgagat tctctcagat
 7500

30 ccaagtgatg ataccaaagg attttttgac cccaacactg aagaaaatct tacctatctg
 7560

caactaaaag aaagatgcat taaggatgag gaaacagggc tctgtcttct gcctctgaaa
 7620

35 gaaaagaaga aacagggtgca gacatcaca aagaataccc tcaggaagcg tagagtggc
 7680

atagttgacc cagaaaccaa taaagaaatg tctgttcagg aggcctacaa gaagggccta
 7740

40 attgattatg aaaccttcaa agaactgtgt gagcaggaat gtgaatggga agaaataacc
 7800

atcacgggat cagatggctc caccaggggtg gtcttggtag atagaaagac aggcagtcag
 7860

45 tatgatattc aagatgctat tgacaagggc cttgttgaca ggaagtctt tgatcagtac
 7920

cgatccggca gcctcagcct cactcaattt gctgacatga tctccttgaa aaatgggtgc
 7980

50 ggcaccagca gcagcatggg cagtgggtgc agcagatgatg ttttagcag ctcccgacat
 8040

gaatcagtaa gtaagatttc caccatatcc agcgtcagga atttaaccat aaggagcagc
 8100

55 tctttttcag acacctgga agaactgagc cccattgcag ccatctttga cacagaaaac
 8160

ctggagaaaa tctccattac agaaggtata gagcggggca tcgttgacag catcacgggt
 9220

5 cagaggcttc tggaggctca ggcctgcaca ygtggcatca tccaccaac cacgggccag
 8280

aagctgtcac ttcaggacgc agtctcccag ggtgtgattg accaagacat ggcaccacg
 8340

10 gtgaagcctg ctcagaaagc cttcataggc ttcgagggtg tgaagggaaa gaagaagatg
 8400

tcagcagcag aggcagtgaag aaaaaatgg ctcccgtatg aggctggcca gcgcttcctg
 8460

15 gagttccagt acctcacggg aggtcttctgt gaccgggaag tgcattgggag gataagcacc
 8520

gaagaagcca tccggaagg gttcatagat ggccgcgccg cacagaggct gcaagacacc
 8580

20 agcagctatg ccaaaatcct gacctgcccc aaaaccaaataaaaaatc ctataaggat
 8640

gccataaatc gctccatggt agaagatatc actgggctgc gccttctgga agccgcctcc
 8700

25 gtgtcgtcca agggcttacc cagcccttac aacatgtctt cggctccggg gtcccgtcc
 8760

ggctcccgt cgggatctcg ctccgatct cgtccgggt cccgcagtgg gtcccgaga
 8820

30 ggaagctttg acgccacagg gaattcttcc tactcttatt cctactcatt tagcagtagt
 8880

tctattgggc actagtagtc agttgggagt ggttgctata ccttgacttc atttatatga
 8940

35 atttccactt tattaataaa tagaaaagaa aatcccgggtg cttgcagtag agtgatagga
 9000

cattctatgc ttacagaaaa tatagccatg attgaaatca aatagtaaag gctgttctgg
 9060

40 ctttttatct tcttagctca tcttaaataa gcagtacact tggatgcagt gcgtctgaag
 9120

tgctaatacag ttgtaacaat agcacaaatc gaacttagga tttgtttctt ctcttctgtg
 9180

45 ttctgatttt tgatcaattc tttaatcttg gaagcctata atacagtttt ctattcttgg
 9240

agataaaaaat taaatggatc actgatattt tagtcattct gcttctcacc taaatatttc
 9300

50 catattctgt attaggagaa aattaccctc ccagcaccag cccccctctc aaaccccaa
 9360

cccaaaacca agcatttttg aatgagtctc cttlagtttc agagtgtgga ttgtataacc
 9420

55 catatactct tcgatgtact tgtttggttt ggtattaatt tgactgtgca tgacagcggc
 9480

aatcttttct ttggtcaaag tttctgttt attttgcttg tcatattcga tgtactttaa
 9540

5 ggtgtcttta tgaagtttgc tattctggca ataaactttt agactttt
 9588

10 <210> 33
 <211> 366
 <212> DNA
 <213> human

<220>
 <221> misc_feature
 <222> (351)..(351)
 15 <223> any kind of base

<400> 33
 gaagtgccat ttatatattat acaaaaatat tacataattc agttaylalt ggtgacataa
 60

20 tttagttagt atgggtgata taatggtcac aatttttagc atctaataaa gatcttttta
 120

tgagtgccat ataaaatatg tgaacaaagc aatcttgta taagatttgt gatgatttag
 180

25 gagaaagtac ttgagataa ttttttctg tctctttgtg aactctctca acagtagttc
 240

tcttttagatt agagccagca ggtcggccat aacagttttc ttcaaatttg ggcaacagtl
 300

30 atacaaatgc ttgaatttca agacaacata ttaaagggtc tatgaactgg naatctaacc
 360

tggggtt
 366

35 <210> 34
 <211> 1466
 <212> DNA
 <213> human

40 <400> 34
 agccccaagc ttaccacctg cacccgaga gctgtgtgtc accatgtggg tcccggttgt
 60

45 ctctctcacc ctgtccgtga cgtggattgg tgtgcaccc ctcactctgt ctccgattgt
 120

gggaggctgg gaggcgaga agcattccca accctggcag gtgcttgtgg cctctcgtgg
 180

50 cagggcagtc tgcggcgggtg ttctggtgca ccccagtggt gtcctcacag ctgcccactg
 240

catcaggaac aaaagcgtga lcttgctggg tcggcacagc ctgtttcatc ctgaagacac
 300

55 aggccaggta ttacaggtca gccacagctt cccacacccg ctctacgata tgagcctcct
 360

EP 1 355 149 A2

gaagaatcga tteclcagge cagglyatga ctccagccac gacctcatgc tgetccgcct
420

5 gtcagagcct gccgagctca cggatgctgt gaaggtnatg gacctgccca cccaggagcc
480

agcactgggg accacctgct acgcctcagg ctggggcagc attgaaccag aggagttctt
540

10 gaccccaaag aaacttcagt gtgtggacct ccatgttatt tccaatgacg tgtgtgcgca
600

agttcacctt cagaagggtga ccaagttcat gctgtgtgct ggacgctgga cagggggcaa
660

15 aagcacctgc tcgggtgatt ctgggggccc acttgtctgt aatgggtgtc ttcaaggtat
720

cacgtcatgg ggcagtgaac catgtgccct gcccgaaagg ccttccctgt acaccaaggt
780

20 ggtgcattac cggaagtgga tcaaggacac catcgtggcc aacctctgag caccctatc
840

aacccctat ttagtaaac ttggaacctt ggaaatgacc aggccaaagc tcaagcctcc
900

25 ccagttctac tgaccttctt ccttaggtgt gaggtccagg gttgctagga aaagaaatca
960

gcagacacag gtgtagacca gagtgtttct taaatgggtg aattttgtcc tctctgtgc
1020

30 ctggggaata ctggccatgc ctggagacat atcactcaat ttctctgagg acacagatag
1080

gatggggtgt ctgtgttatt tgtgggttac agagatgaaa gaggggtggg atccacactg
1140

35 agagagtgga gagtgcacg tgctggacac tgtccatgaa gcactgagca gaagctggag
1200

gcacaacgca ccagacactc acagcaagga tggagctgaa aacataaccc actctgtcct
1260

40 ggaggcactg ggaagcctag agaaggctgt gagccaagga gggagggtct tcctttggca
1320

tgggatgggg atgaagtaag gagagggact ggacccctg gaagctgatt cactatgggg
1380

45 ggaggtgtat tgaagtctc cagacaaccc tcagatttga tgatttccta gtagaactca
1440

cagaaataaa gagctgttat actgtg
1466

50 <210> 35
<211> 187
<212> DNA
<213> human

<400> 35
55 gatctggcgc attccggctg acactctcgt ttatttggac tglaagtctg acctctatga
60

EP 1 355 149 A2

ataattantt cagccccctga gtgctcccgg gccaaagctcc ttggccaaac tttcacctta
120

5 gctttctgata agtcttgggc caagctaagc agcatctatc aatcatccct tcagctcctg
180

attgac
187

10

<210> 36
<211> 2913
<212> DNA
<213> human

15 <400> 36
actgggtacc gaggactggg tgtgtttaaq gcagacagcc aggtgaggat cccagctact
60

ggggcctgct gtcactcctt gggagtaccc gggggtcagg agcctagggg actcttgac
120

20 ttcacatcca gccatgctaa ttacactttt tggcaaagga aacagctagg agcagtttct
180

ttacttcta cagccccgtt ttctcagtgt ttagacctcg aattattact gggctagagg
240

25 gaaggcagcc tctgaagtgt ggcaggagga ggggaagtct gcctgcatct tgggtgtgct
300

gtcagatgcc agcactaata acctggcttc tgtgaggcct gtcagtgtct tcaggaatga
360

30 aaggggaccc ctgagaggtg ctcagtagca gcaggtgtg aatgctctct acccaccacc
420

ctcacctcct cgtaaagat ggtgtacct gccacacagc agacatctgg tcgctgcaca
480

35 cccgaagac cccaaggcag tctgccccct gtccagccac acgccagcac ccacctcct
540

ggccccctgcc tcggcctccc cagaccagct gcacccagcc cccaacacgc accccttctc
600

40 cagatgtgtg cagggcctca ttttgagag caaagacaga tgtttcagcc acacgcttta
660

ttaacttcta aaacctgtgc tcaggacact cttcaacagt catgaaaagt ttgatcactt
720

45 gccacagtca ggacctttgt gtggggctct gatctgatgt tcggctcat catctcccaa
780

accagcagtc gtttgtacct caacctctg ctgaggggt cataccccca aatgattttc
840

50 ctgatttatg tatttcctta caaagggtt tctataccta gcctctgcct ccagcatgag
900

aagggggaat aggtgagacc catttgccag tagcagacgg ggacctggg gagaaaatgg
960

55 cagagcctgt tggagactcc ctgtctccag ctgaccagcc aatgggattc ctcttccctc

1020

cactgtctcc cacaaagtag aagaatcctg gtacatttag cccatgagcc tggcacagat
1080

5

ccctatctag acatgaggcc ctttagacat gactttggca ttgaccagcc tgttggcaat
1140

gggtcgggga ggcagagggg atgctcacac cagtaattct catccctga atgcttggga
1200

10

tcacctgggg agagttcaca aaatactggt gcaggggtcc cacctctgat gatgctgagt
1260

ggtgggtctg ggtgtggcc caggcatcat gatgtttcag gccccaggt gacttcttag
1320

15

gcagcccagc taagccccta gagccttgca atttcccca aatgacctca gagggccga
1380

tttgagggaa atgcctaact tcaggggccc taagaatccc ccaggagca tgtgaaatgc
1440

20

agataccagg cccaccccca gagatgagct gaggtgggtc aggggtgaag tgcagggatc
1500

agtgttttc acaagctcca tacctccagg aaatggtgtt gtggttgggc ccgtagaaaa
1560

25

cattctgaga gtccgtgtgc ctgtgccttg gtgcacgtgg ggtggaatcc cagtggccct
1620

gccttgagga ggatgtgcat taacgtggtg ggggagacag agacagctcc acctgcccc
1680

30

tgtcccaccg gggacctcca aaaacttcat ggatgttaga gcaagcagcc atgctgcagc
1740

agaggatgag gctggcggat ttagtaagag cctctgtgt ttgggctgag ttctttctct
1800

35

agttgcctg tcatctggcc tctggataac ccacctctcc tccctcatcc taaaattaca
1860

gatggcgaaa gacggccaca tttagtgaga cccctaaggt cctccaacta ggggtgggtcc
1920

40

acagtggccc ctggtgcatg gaccacacac tctcttcct cctctggctc aggactacgg
1980

tctgaaatta gggagatatg aatgtcttcc ttgaaaactt ctcttcccag tcttcccact
2040

45

ttgcttgggg gtccttggtc aaggccagct ttggactagg gcttgttgcg actaccagct
2100

gtctcatttt gctgtactgc aaactcaggc ttggttccaa gcttatgggg gccctgtcct
2160

50

tcccctagta ggglttgttt tggggtcaca tctggtcata cccttcagag agctcttccc
2220

cagcctctac atcagggaga gaggtaggta gggaggagca ttcaaggatt agaagaagga
2280

55

ctaaagtaca acagccttgg aggaactgcc aggaactaag ggcgagcact ggagaaggca

2340

acctgggacc ccctgcgctt ctgagcagga agaccaagac cttcaggggc cctaagcact
2400

5

gaaaacatca ttctcatcc ccaagccctg gcatccccct gttcttctaa aataattctt
2460

ttctaggtat ttctgattgc aaaattctgg atgggttcat ccaagctgac ctttctgtt
2520

10

ttttcccttc ccaacaaggc ctcacttttt ggagccacct tagttggtgc ctaggcagag
2580

gggcagtcag cagtggltat caggatcctg gctctatggg ttgccttctt cctgggtctgt
2640

15

aaagcccctg caggcaggga cttcttagat agctgcttcc ttagggcatg gcatgtggtg
2700

ggtggttaat gaatggaaga gagggaaatga gtgatcaagg gagggaggag ggagtggagt
2760

20

ggagatttct catcctttcc tgttaattta tgacatcctc ctgcctatga gtccttgact
2820

ctggagtttt acaaagcagt cacatttcaa alaaaagtct gggaaagcaa cacatcatcg
2880

25

ccaactttta attttgctaa ataaggatat tag
2913

<210> 37
<211> 1466
<212> DNA
<213> human

30

<400> 37
agccccaagc ttaccacctg caccgggaga gctgtgtgtc accatgtggg tcccggttgt
60

35

cttcctcacc ctgtccgtga cgtggattgg tgctgcaccc ctcacctgt ctcggattgt
120

gggaggctgg gagtgcgaga agcattccca accctggcag gtgcttgtgg cctctcgtgg
180

40

cagggcagtc tgcggcggtg ttctggtgca ccccagtggt gtccctcacag ctgcccactg
240

catcaggaac aaaagcgtga tcttgcctggg tcggcacagc ctgtttcatc ctgaagacac
300

45

aggccaggta ttacaggtea gccacagctt cccacacccg ctctacgata tgagcctcct
360

gaagaatcga ttctcaggc cagggtgatga ctccagccac gacctcatgc tgctccgcct
420

50

gtcagagcct gccgagctca cggatgctgt gaaggtcatg gacctgccca cccaggagcc
480

agcactyggg accacctgct acgcctcagg ctggggcagc attgaaccag aggagttctt
540

55

EP 1 355 149 A2

gaccccaaaag aaacttcayt gtgtggacct ccatgttatt tccaatgacg tgtgtgcgca
600

5 agttcaccct cagaaggatga ccaagttcat gctgtgtgct ggacgctgga caggggggcaa
660

aagcacctgc tcgggtgatt ctggggggccc acttgtctgt aatgggtgac ttcaaggtat
720

10 cacgtcatgg ggcagtgaac catgtgccct gcccgaaagg ccttccctgt acaccaaggt
780

ggtgcattac cggaagtgga tcaaggacac catcgtggcc aaccctgag caccctatc
840

15 aacccttat ttagtagaac ttggaacctt ggaaatgacc aggccaagac tcaagcctcc
900

ccagttctac tgacctttgt ccttaggtgt gaggtccagg gttgctagga aaagaaatca
960

20 gcagacacag gtgtagacca gagtgtttct taaatggtgt aattttgtcc tctctgtgac
1020

ctggggaata ctggccatgc ctggagacat atcactcaat ttctctgagg acacagatag
1080

25 gatggggtgt ctgtgttatt tgtgggttac agagatgaaa gaggggtggg atccacactg
1140

agagagtgga gagtgcacat tgctggacac tgtccatgaa gcactgagca gaagctggag
1200

30 gcacaacgca ccagacactc acagcaagga tggagctgaa aacataaccc actctgtcct
1260

ggaggcactg ggaagcctag agaaggctgt gagccaagga gggaggggtct tcctttggca
1320

35 tgggatgggg atgaagtaag gagagggact ggacccctg gaagctgatt cactalgggg
1380

ggaggtgtat tgaagtcctc cagacaaccc tcagatttga tgatttccta gtagaactca
1440

cagaaataaa gagctgttat actgtg
1466

40

<210> 38
<211> 462
<212> DNA
<213> human

45

<220>
<221> misc_feature
<222> (197)..(197)
<223> any kind of base

50

<220>
<221> misc_feature
<222> (116)..(116)
<223> any kind of base

55

5 <220>
 <221> misc_feature
 <222> (334)..(334)
 <223> any kind of base

10 <220>
 <221> misc_feature
 <222> (402)..(402)
 <223> any kind of base

15 <220>
 <221> misc_feature
 <222> (429)..(429)
 <223> any kind of base

20 <220>
 <221> misc_feature
 <222> (438)..(438)
 <223> any kind of base

25 <220>
 <221> misc_feature
 <222> (443)..(443)
 <223> any kind of base

30 <220>
 <221> misc_feature
 <222> (459)..(459)
 <223> any kind of base

35 <400> 38
 taagggtttta taattatttt tattttttctt ttcttttttt tttatggctt ggatgacact
 60
 ttattttcag atccaatact agaagttggt tccatgttca cattttcctt cctggnttaa
 120
 aaaaaagagt tgtatttttt ttttttgctt tttttaaat atactttaag ttttagggta
 180
 catgtgcaca acgcagnngt tagctacata tgtatacatg tgccatggtg gcgtgctgca
 240
 tccagtaact cgtcatttaa cattagglat atctccaaat gctatccttc ccccatgtgt
 300
 atttttcata gcttaaaaaat cattgacata ggantaattc caactaaagt acggtattaa
 360
 atccctgggg gaataaattt tgtcttaaca agggtaaggt tngtgaaaag gatggttttg
 420
 tcacagggna aaaggganat ccncccatth taaaaccenc ct
 462

50 <210> 39
 <211> 1490
 <212> DNA
 <213> human

55

EP 1 355 149 A2

<400> 39
ctcgtgcccc ccacggaggg gactgctctc ccccgtgca tcctttctgt gaggtacctt
60

5 acccacctca gcacctgaga gggtgaaata gaattctaac ctcgacattc gggaagtgtt
120

tttgagaagt ctcggtcggg aagggagtc ttccaagtc gtgcagcact aacgtattgg
180

10 cacctgcctc ctcttcggcc acccccaga tgaggcagct gtgactgtgt caaggggaagc
240

cacgactctg accatagtct tctctcagct tccactgccg tctccacagg aaaccacaga
300

15 gttctgtgaa caagtccatg ctgccatcaa ggcattttatt gcagtgtact atttgcttcc
360

aaaggatcag gggatcaccc tgagaaagct ggtacggggc gccaccctgg acatcgtgga
420

20 tggcatggct cagctcatgg aagtactttc cgtcactcca actcagagcc ctgagaacaa
480

tgaccttatt tctacaaca gtgtctgggt tgcgtgccag cagatgcctc agataccaag
540

25 agataacaaa gctgcagctc ttttgatgct gaccaagaat gtggattttg tgaaggatgc
600

acatgaagaa atggagcagg ctgtggaaga atgtgacct tactctggcc tcttgaatga
660

30 tactgaggag aacaactctg acaaccacaa tcatgaggat gatgtgttgg gggttcccag
720

caatcaggac ttgtattggg cagaggacga tcaagagctc ataatcccat gccttgcgct
780

35 ggtgagagca tccaaagcct gcctgaagaa aattcggatg ttagtggcag agaatgggaa
840

gaaggatcag gtggcacagc tggatgacat tgtggatatt tctgatgaaa tcagccctag
900

40 tgtggatgat ttggctctga gcataatcc acctatgtgt cacctgaccg tgcgaatcaa
960

ttctgcgaaa cttgtatctg ttttaagaa ggcacttgaa attacaaaag caagtcattg
1020

45 gacctctcag ccagaagata gttggatccc ttactttatt aatgccattg atcattgcat
1080

gaatagaatc aaggagctca ctcagagtga acttgaatta tgacttttca ggetcatltg
1140

50 tactctcttc ccctctcctc gtcattgtca ggctctgata cctgctttta aaatggagct
1200

agaatgcttg ctggattgaa agggagtgc tctctatatt tagcaagaga cactattacc
1260

55 aaagattgll ggtaggcca gattgacacc tatttataaa ccatatgcgt atatttttct

1320

gtgctatata tgaaaaataa ttgcatgatt tctcattcct qagtcatttc tcagagattc
1380

ctaggaaagc tgccttattc tclttttgca gtaaagtatg ttgttttcat tgtaaagatg
1440

ttgatggtct caataaaatg ctaacttgcc agtgattaaa aaaaaaaaaa
1490

<210> 40
<211> 1677
<212> DNA
<213> human

<400> 40
cttgacccta tttatagtgg ctctaaagggt ggtgttatta tgttttctag agcacttcga
60

ttatacaaac gtcaaggaat ccgagttaat gtgctttgcc ctgagtttgt tgaaacagac
120

atgggcacaa tgatcggtcc caaattcctt agtatgatgg ggggctttgt acctatggaa
180

atggtggtga aagggtgcttt tgagctcatc actgatgaga ataaagccgg cgattgccta
240

tggattacta atcggcgagg tcttgagtac tggcccaccc catcagaaga agcaaagtac
300

ttgctgcgtt ctacacgttc caggagaaga actgaataca aagctccacc aattaaacta
360

cctgagagtt ttgagaaaat agttgttcag accttgactc acaactttcg gaatgctacc
420

agtgtagtaa gagcaccact gagattacct atcaaaccac actatgttct tgtgaagata
480

atctatgctg gtgtaaatgc tagtgatgta aattttagct caggtcgcta ttttgggtggc
540

aataacagtg aactgcatc ccgtcttcgg tttgatgcag gatttgaggc tgtgggagta
600

attgcagcag ttggggattc tgttactgac ttgaaagttg gcatgccttg tgcgttcacg
660

acttttggag gctatgctga atttacaatg attccttcga aatacgccct tccaatgcct
720

agaccagaac cggaaggtyt tgccatgctt acatcaggat taacagcttc aattgctcta
780

gaaaaggcag gacagatgga atctggaaaa gtggtccttg ttactgctgc ggcaggagga
840

actggtcagt ttgctgttca gcttgcaaaa ttagctggta ataccgtggt tgccacttgt
900

ggaggtgggg caaaggccaa gcttctgaaa gaattgggag tcgacagagt catagactat
960

EP 1 355 149 A2

cacagtgaag atataaaaac ggttctaagg aaagagttcc cgaaagggtat tgatatcatc
 1020
 5 lacgaatctg ttggtgggga catgttaaag ttgtgcttgg atgctttggc agtccatgga
 1080
 cgactcattg tcattggcat gatttctcag latcaaggag aaaatgggtg gacgccatca
 1140
 10 aaatatcctg gactatgtga gaagctcttg tcaaagagtc aaactgtggc tggctttttc
 1200
 ctggtgcaat atagtacat gtaccaagaa caccttaaca agttatttga cctttactct
 1260
 15 tccggaaaaac taaaggttgc tgtggatcca aagagattta taggccttca ttctgttgct
 1320
 gatgctgttg agtatctcca ttcaggcaaa agcgttggga aggtgggtgt ctgctgggac
 1380
 20 ccgaccttcg gicatcaagt agccaaatta tgaatgaaca cgggtgtcaaa tacagaaaga
 1440
 agtgaagttt tcaattctta gtctagagat tgttctcgaa tgttactgaa aatagctgct
 1500
 25 agaccagtgc tggaatattt attctcaatg ctttttcaat tttggattac ttgaaagaat
 1560
 aatccattta tgtataccat gtttatgttt acactataca acaactatga gcagaagaaa
 1620
 30 gcgagatata taaaaataa attataatcc tttcatttta aaaaaaaaaa aaaaaaa
 1677
 <210> 41
 <211> 1330
 <212> DNA
 35 <213> human
 <400> 41
 atggcgcagt gggacagctt cactgatcaa caggaggaca ctgatagctg ttcagaatct
 60
 40 gtgaagtttg atgctcgctc caatacagct ttgcttcccc caaatcctaa aaatggccct
 120
 ccacttcaag aaaagctgaa atccttcaaa gctgcactga ttgcccttta tctccttgctg
 180
 45 tttgctgttc tcattcctat catcgcaata atggcagctc aactcctgaa gtgggaaatg
 240
 aagaattgca cagttgggtc aattaatgca aacagtgtat cctccagtct cctgggaaga
 300
 50 ggaaatgaca gtgaagatga agtgagattt cgagaagttg ttatggaaca cattagcaag
 360
 atggagaaaa gaatccaata tatttcagat actgaagaaa atctcgtaga ttcagagcat
 420
 55 tttcaaaatt tcagtgtgac aactgatcaa cgatttgctg atgttcttct ccaactaagt
 480

accttgggtc ccacagtcca gggacatggg aatgccgtag atgaaatcac caggtcctta
540

5 ataagtctga ataccacgct gcttgatttg cacctctatg tagaaacact gaatgtcaaa
600

ttccaggaga atacacttaa agggcaagag gaaatcagca aattaaagga gcgtgtgcac
660

10 aatgcatcag cagaaattat gtctatgaaa gaagaacaag tgcatttggg acaggaaata
720

aaaagagaag tgaaagtcct gaataacatc actaatgatc tcaggctgaa agattgggaa
780

15 cattctcaga cgttgagaaa tatcacttta attcaaggtc ctcttgacc cccaggagaa
840

aaaggagata gaggtccaac cggagaaagt ggtccaccag gcgttcacagg tccagtaggt
900

20 cctccaggtc tlaagggatga tcgaggatct attggctttc cgggaagtcg aggatatcca
960

ggacaatcag ggaagactgg gaggacagga tatcctggac caaaaggcca aaagggaqaa
1020

25 aaaggcagtg gaagcatcct gactccttct gcgactgtcc gactggttgg lggccgtggc
1080

cctcatgagg gtagagtggg gataatgcac aatggacagt ggggcacagt ttgtgatgat
1140

30 cactgggaac tgcgtgccgg gcaggttgtc tgcaggagct tgggataccg aggtgttaag
1200

agtgtgcaca agaaagctta ttttggacaa ggtactggtc ccatttggct gaatgaagta
1260

35 ccctgtttgg gtagggagtc atccattgaa gagtgcacaa tcagacagtg gggcgtgaga
1320

gtctgttcac
1330

40

<210> 42
<211> 431
<212> DNA
<213> human

45

<220>
<221> misc_feature
<222> (97)..(97)
<223> any kind of base

50

<220>
<221> misc_feature
<222> (347)..(347)
<223> any kind of base

55

<220>
<221> misc_feature

<222> (349)..(349)
 <223> any kind of base

5 <220>
 <221> misc_feature
 <222> (361)..(361)
 <223> any kind of base

10 <220>
 <221> misc_feature
 <222> (362)..(362)
 <223> any kind of base

15 <220>
 <221> misc_feature
 <222> (363)..(363)
 <223> any kind of base

20 <220>
 <221> misc_feature
 <222> (401)..(401)
 <223> any kind of base

25 <220>
 <221> misc_feature
 <222> (428)..(428)
 <223> any kind of base

30 <400> 42
 ctttttatat ttattttcat cgctacacaa acatttttta ggagtttgat tctacctcca
 60

 ttttggttag atatacaaac tctaccccat gagggantgt atggtgtatt tctagattta
 120

35 gcaacaattt tcttgaaaaa tgtacaatac tatagaaaaa tgaagatagt aaataccagg
 180

 tataagttaa taacagtgtt tcttttgttc agtaataatg aactgtgtac tagcactgaa
 240

40 ctttaggccc tcctatttgc gtattttctg tttgtatatt tttaaataga ggaattgtga
 300

 ttataatatt attattttgg aatatacctaa atcataaatt caaaacntna tttagttttt
 360

45 nnnttttttt tttaagatgg agtcccgtt tgtcccaggc nggagtgcag tggcatgac
 420

 tcagctcnc t g
 431

50 <210> 43
 <211> 669
 <212> DNA
 <213> human

55 <220>

EP 1 355 149 A2

<221> misc_feature
<222> (641)..(641)
<223> any kind of base

5

<400> 43
ttcttttggg aaaccaaaca tgctttatct catttttttc acaatttatt taaacatctc
60

10

acatatacaa aataggtaca atttaatttt tctgcttgcc caagaaacaa agcttctgtg
120

gaaccatgga agaagatgaa aatgagactg gcaaagaaca aatgctgaat ctgaagaaga
180

15

ggacaacttt gggcaaataa tctgcatact tttaattggg aataagatgg aaaatatgaa
240

tgctaaatca aattttttta aaaaatacacc acacgataca actcaatata ggagtatttc
300

20

ttctcaaatt cttctagcac catcaacatt cttcaagtat ctgaaatact attaattagc
360

acctttgtat tatgaacaaa acaaaacaag gacctcagtt catcctgtgc taggtcagca
420

25

cctaacaatg tggatcacac tcatgggaaa gtgttttgag gtagtttaaa cctttggaag
480

tttgggtttt aaacttcctt ctgtggaaga tattcaaaag ccacaagtgg tgcaaatgtt
540

30

tatgggtttt atttttcaat ttttattttg gttttcttac aaagggtgac atttttcata
600

acagggtgta gagtggtgaa aaaaaaattt caattttttg ngggaacggg ggaaggagtt
660

35

aatgaaact
669

40

<210> 44
<211> 287
<212> DNA
<213> human

<400> 44
gccggagagt ctacaatgtt acccagcatg ctgttggcat tgttgtaaac aaacaagtta
60

45

agggcaagat tcttgccaag agaattaatg tgcgtattga gcacattaag cactctaaga
120

gccgagatag cttcctgaaa cgtgtgaagg aaaatgatca gaaaaagaaa gaagccaaag
180

50

agaaagggtac ctgggttcaa ctaaagcgcc acgctgctcc acccagagaa gcacactttg
240

tgagaaccaa tgggaaggag cctgagctgc tggaacctat tccctat
287

55

<210> 45

<211> 383
 <212> DNA
 <213> human

5 <220>
 <221> misc_feature
 <222> (147)..(147)
 <223> any kind of base

10 <220>
 <221> misc_feature
 <222> (309)..(309)
 <223> any kind of base

15 <220>
 <221> misc_feature
 <222> (349)..(349)
 <223> any kind of base

20 <220>
 <221> misc_feature
 <222> (365)..(365)
 <223> any kind of base

25 <220>
 <221> misc_feature
 <222> (372)..(372)
 <223> any kind of base

30 <220>
 <221> misc_feature
 <222> (380)..(380)
 <223> any kind of base

35 <400> 45
 ggaacggaaa aggagaattc aagtgtgacc ctcatgaggc aacgtgttat gatgatggga
 60
 agacatacca cgtaggagaa cagtggcaga aggaatatct cggtgccatt tgctcctgca
 120

40 catgctttgg aggccagcgg ggctcgnctt gtgacaactg ccgcagacct ggggggtgaa
 180
 cccagtcctg aaggcactac tggccagtcc tacaaccagt attcttcaga gataccattc
 240

45 agagaacaaa cactaatggt taatttgccc aatttgagtg ctcatgcct tttaggatgt
 300
 tacaggctng acagagaagg ttttcccgag gagttaaate atctttttnc catttccga
 360

50 ggggncagg cntgttttn ttt
 383

55 <210> 46
 <211> 523
 <212> DNA

EP 1 355 149 A2

<213> human

<400> 46
5 cagaggggca gggcgacgg ctaggagttc aagaacatc ctggtctgag ggaaaggctg
60
cagctgcacc gccatgaata agcttttcag cttctggaag aggaagaatg agacccgcag
120
ccagggtac aaccttcgag aaaaggattt aaagaaactt cacagagctg cttcagtcgg
10 180
ggatttgaag aagctgaagg aalaccttca gatcaagaaa tatgatgtaa atatgcagga
240
ctatgaatac agaacacctt tgcacctagc ctgtgctaata ggacatacag atgttggtact
15 300
tttcctaatt gagcaacaat gcaagataaa tgtccgggat agtgaaaaca aatccccatt
360
gattaaggca gtacagtgtc aaaatgagga ttgtgtctact attctgctaa actttgggtgc
20 420
agaccagat ctgagggata ttcgttataa tactgttctt cactatgctg tttgtggtca
480
aagtttgtca ttagttgaaa aactgcttga atacgaagct gat
25 523

<210> 47
<211> 390
<212> DNA
30 <213> human

<400> 47
tccaagggtca tggcaaaaca tctgaagttc atcgccagga ctgtgatggt acaggaaggg
60
aacgtggaaa gcgcatacag gaccctaaac agaatcctca ctatggatgg gctcattgag
35 120
gacattaagc atcggcggta ttatgagaag ccatgccgcc gcgacagagg gaaagctatg
180
aaaggtgccg gcggatctac aacatggaaa tggctcgcaa gatcaacttc ttgatgcgaa
40 240
agaatcgggc agatccgtgg cagggtgct gaggcctgtg ggtgggacac cagtgcgaaa
300
ccctcatcca gtttctctc catctctttt ctttgtacaa tcccatttcc tattaccatt
45 360
ctctgcaata aactcaaac acatgtctgc
390

50 <210> 48
<211> 669
<212> DNA
<213> human

55 <220>
<221> misc_feature

<222> (641)..(641)
 <223> any kind of base

5 <400> 48
 ttcttttgga aaaccaaacca tgctttatctt catttttttc acaatttatt taaacatctc
 60
 acatatacaa aataggtaca atttaatttt tctgcttgcc caagaaacaa agcttctgtg
 120
 10 gaaccatgga agaagatgaa aatgagactg gcaaagaaca aatgctgaat ctgaagaaga
 180
 ggacaacttt gggcaaataa tctgcatact tttaattggg aataagatgg aaaatatgaa
 240
 15 tgctaaatca aatttttttaaaa aaatacacc acacgatata actcaatata ggagtatctc
 300
 ttctcaaatt cttctagcac catcaacatt cttcaagtat ctgaaatact attaattagc
 360
 20 acctttgtat tatgaacaaa acaaaacaag gacctcagtt catctctgtc taggtcagca
 420
 cctaacaatg tggatcacac tcatgggaaa gtgttttgag gtagtttaaa cctttggaag
 480
 25 tttgggtttt aaacttccct ctgtggaaga tattcaaaag ccacaagtgg tgcaaatgtt
 540
 tatgggtttt atttttcaat ttttattttg gttttcttac aaagggtgac atttttcata
 600
 30 acagggtgta gagtggtgaa aaaaaaattt caatttttgg ngggaacggg ggaaggagtt
 660
 aatgaaact
 669
 35
 <210> 49
 <211> 431
 <212> DNA
 <213> human
 40
 <220>
 <221> misc_feature
 <222> (97)..(97)
 <223> any kind of base
 45
 <220>
 <221> misc_feature
 <222> (347)..(347)
 <223> any kind of base
 50
 <220>
 <221> misc_feature
 <222> (349)..(349)
 <223> any kind of base
 55
 <220>

<221> misc_feature
 <222> (361)..(361)
 <223> any kind of base

5

<220>
 <221> misc_feature
 <222> (362)..(362)
 <223> any kind of base

10

<220>
 <221> misc_feature
 <222> (363)..(363)
 <223> any kind of base

15

<220>
 <221> misc_feature
 <222> (401)..(401)
 <223> any kind of base

20

<220>
 <221> misc_feature
 <222> (428)..(428)
 <223> any kind of base

25

<400> 49
 ctttttatat ttattttcat cgctacacaa acatttttta ggagtttgat tctacctcca
 60

30

ttttggtttag atatacaaac tctaccccat gagggantgt atggtgtatt tctagattta
 120

gcacaattt tcttgaaaaa tgtacaatac tatagaaaaa tgaagatagt aaataccagg
 180

35

tataagttaa taacagtgtt tcttttggtc agtaataatg aactgtgtac tagcactgaa
 240

ctttaggccc tcctatttgc gtattttctg tttgtatatt tttaaataga ggaattgtga
 300

40

ttataatatt attatttttg aatattcctaa atcataaatt caaaacntna tttagttttt
 360

nnnttttttt ttttaagatgg agtcccgtt tgtcccaggc nggagtgcag tggcatgatc
 420

45

tcagctcnct g
 431

50

Claims

55

1. A method of assessing colorectal cancer status comprising identifying differential modulation of each gene (relative to the expression of the same genes in a normal population) in a combination of genes selected from the group consisting of Seq. ID. No. 42-49.

2. The method of claim 1 wherein there is at least a 2 fold difference in the expression of the modulated genes.
3. The method of claim 1 wherein the p-value indicating differential modulation is less than .05.
- 5 4. The method of claim 1 further comprising employing a colorectal diagnostic that is not genetically based.
5. The method of claim wherein the cancer marker that is not genetically based is selected from the group consisting of carcinoembryonic antigen, CA19-9, CA 125, CK-BB, and Guanylyl Cyclase C.
- 10 6. A diagnostic portfolio comprising isolated nucleic acid sequences, their complements, or portions thereof of a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
7. The diagnostic portfolio of claim 6 in a matrix suitable for identifying the differential expression of the genes contained therein.
- 15 8. The diagnostic portfolio of claim 7 wherein said matrix is employed in a microarray.
9. The diagnostic portfolio of claim 8 wherein said microarray is a cDNA microarray.
- 20 10. The diagnostic portfolio of claim 8 wherein said microarray is an oligonucleotide microarray.
11. A diagnostic portfolio comprising isolated nucleic acid sequences, their complements, or portions thereof of a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
- 25 12. A kit for diagnosing colorectal cancer comprising isolated nucleic acid sequences, their compliments, or portions thereof of a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
13. The kit of claim 12 further comprising reagents for conducting a microarray analysis.
- 30 14. The kit of claim 12 further comprising a medium through which said nucleic acid sequences, their complements, or portions thereof are assayed.
15. A method of assessing response to treatment for colorectal cancer comprising identifying differential modulation of each gene (relative to the expression of the same genes in a normal population) in a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
- 35 16. The method of claim 15 wherein the assessment of the response to therapy includes a determination of whether the patient is improving, not improving, relapsing, likely to improve, or likely to relapse.
- 40 17. Articles for assessing colorectal cancer status comprising isolated nucleic acid sequences, their complements, or portions thereof of a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
18. Articles for assessing colorectal cancer status comprising representations of isolated nucleic acid sequences, their complements, or portions thereof of a combination of genes selected from the group consisting of Seq. ID. No. 42-49.
- 45

50

55